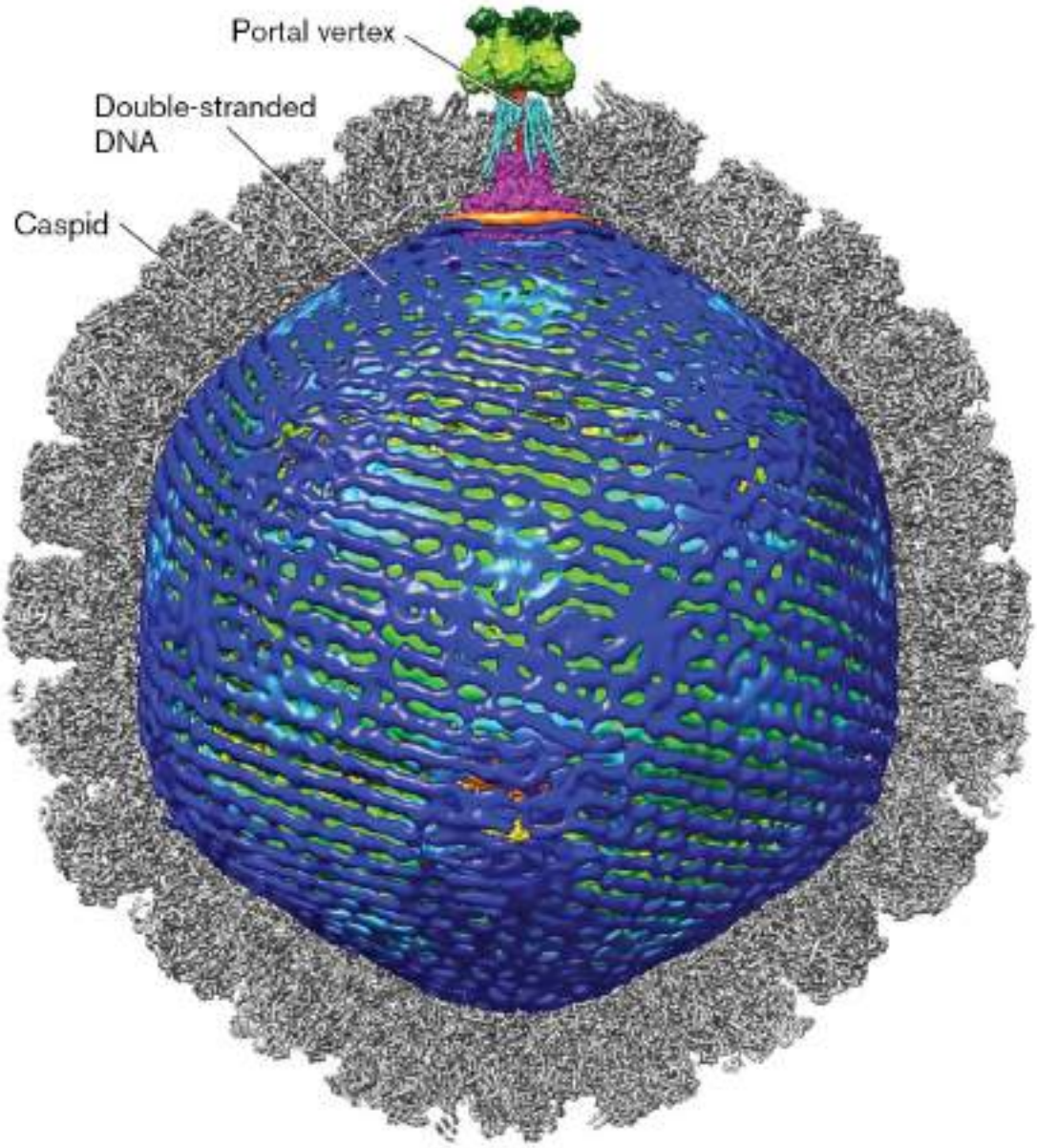
CHAPTER 11

Viral Molecular Biology



Herpes simplex virus is imaged by cryo-electron microscopy (cryo-EM). The viral DNA (colorized blue) is spooled within the protein capsid (gray, section view). The capsid's portal vertex (pink and green) inserts viral DNA into the nucleus of the host cell (not shown).

Y.-T. LIU ET AL. 2019. *NATURE* **570** :257–261

Chapter Sections		Learning Objectives
11.1 Phage Lambda: Enteric Bacteriophage	»	Describe the structure of phage lambda, and explain how its molecular switch chooses lysis or lysogeny.
11.2 Influenza Virus: (–) strand RNA Virus	»	Describe the infectious replication cycle of human influenza viruses and explain the emergence of pandemic mutants.
Special Topic 11 Designing a Pandemic Flu		
11.3 Human Immunodeficiency Virus	»	Describe the replication cycle of human immunodeficiency virus

Chapter Sections		Learning Objectives
(HIV): Retrovirus		(HIV), and explain the formation of quasispecies.
11.4 Endogenous Retroviruses and Lentiviral Gene Therapy	»	Explain how endogenous retroviruses arise in the human genome, and describe how lentivectors derived from HIV provide lifesaving therapies for cancer and genetic defects.
11.5 Herpes Simplex Virus: DNA Virus	»	Describe the replication cycle of herpes simplex virus, and explain the basis of latent infection.
eResearch Activity 11 Could a Lentiviral Vector with CRISPR Cure Herpes?		

As we learned in Chapter 6 , the coronavirus that caused the COVID-19 pandemic is surprisingly simple in form: an envelope with glycoproteins, matrix proteins, and a single-stranded RNA genome that specifies just a few kinds of proteins. How does such a simple reproductive entity propagate throughout the body? Molecular interactions, such as a viral glycoprotein binding a cell-surface receptor, enable a simple virus to take over the host cell's machinery and thwart its defenses. For instance, consider another RNA virus,

the influenza virus, whose seasonal strains kill up to half a million people each year. Influenza viruses show extraordinary diversity of envelope proteins. The structure of these cell-surface proteins determines whether a strain of influenza virus can bind and succeed in infecting a given human or animal host. Once inside a host cell, the viral genome and packaged molecules must collaborate with host enzymes to build virions and to evade defense molecules of the immune system.

Yet our bodies and our environment are also full of viruses that coexist with us without harm, including bacteriophages of our gut bacteria as well as herpes-related viruses. What makes one virus deadly and another helpful? The answer lies in viral molecules that control replication and latency. The functional molecules of influenza and herpes provide targets for new antiviral therapies. Lysogenic phages carry genes among bacteria, whereas lentiviruses alter the genomes of human cells. Endogenous viral genomes contribute molecular parts for our bodies, such as a protein for placental fusion. Even the viruses most deadly to humans can be converted to vectors that cure a human disease.

Chapter 11 also explores three important viruses that infect humans, and whose biology has been studied extensively:

- **Influenza A virus (IAV), a negative-strand RNA virus.** Influenza viruses have caused periodic pandemics, often resulting from genome segment reassortment between human and animal strains.
- **Human immunodeficiency virus (HIV), a type of retrovirus called a lentivirus.** HIV copies its RNA genome into DNA and integrates its genome into that of the host cell, causing lifelong infections worldwide. The lentiviral integration capability allows construction of vectors for gene therapy.
- **Herpes simplex, a large double-stranded DNA virus.** The virus causes skin outbreaks, then undergoes latent infection in nerve cells. Related viruses cause chickenpox

(herpes zoster) and infectious mononucleosis (Epstein-Barr virus; EBV). Herpesviruses are engineered to destroy tumors.

The molecular details of additional viruses are presented in eAppendix 4 :

- **Phage T4 of *Escherichia coli* .** Phage T4 has a relatively large genome encoding an intricate capsid whose assembly requires a large number of steps. The self-assembly of phage T4 provided an early model for study of animal development.
- **Phage M13 of *E. coli* .** Phage M13 replicates by slow release, neither lytic nor lysogenic. Its flexible filamentous length makes it useful for phage display engineering of vaccines and nanomachine components, earning the 2018 Nobel Prize in Chemistry for George Smith and Gregory Winter.
- **Human poliovirus.** Poliovirus is a plus-strand RNA enterovirus that infects the gut and may cause the paralytic disease poliomyelitis (polio). While vaccination has eliminated polio in most countries, related enteroviruses occasionally lead to cases of poliolike paralysis.
- **Human hepatitis C virus (HCV).** A plus-strand RNA virus, HCV propagates in the liver and can cause cancer requiring liver transplant. The HCV replication cycle shares features with those of poliovirus and coronavirus, and it generates a quasispecies similar to that of HIV.

11.1 Phage Lambda: Enteric Bacteriophage



Bacteriophages, or phages—the viruses that infect bacteria—are among the most abundant life forms on Earth. Their ability to infect human pathogens has brought renewed interest for human therapy (as presented in eResearch Activity 11), and the development of phages for therapy requires the ability to engineer their molecular parts. Much of what we know of molecular virology began with the model bacteriophage lambda (**Fig. 11.1**). Bacteriophage lambda infects *Escherichia coli* within the human gut, among the trillions of phages in our intestinal microbiome. Lambda is a "tailed" phage that inserts its double-stranded DNA genome into the bacterial cytoplasm. The phage genome then makes a molecular "choice" between lysis and lysogeny (introduced in Chapter 6). In microbial communities such as our gut, lysogeny provides a way for many kinds of phages to transfer genes between the genomes of diverse bacteria. Phage lysogeny plays vital roles in all microbial communities of humans and environmental ecosystems.



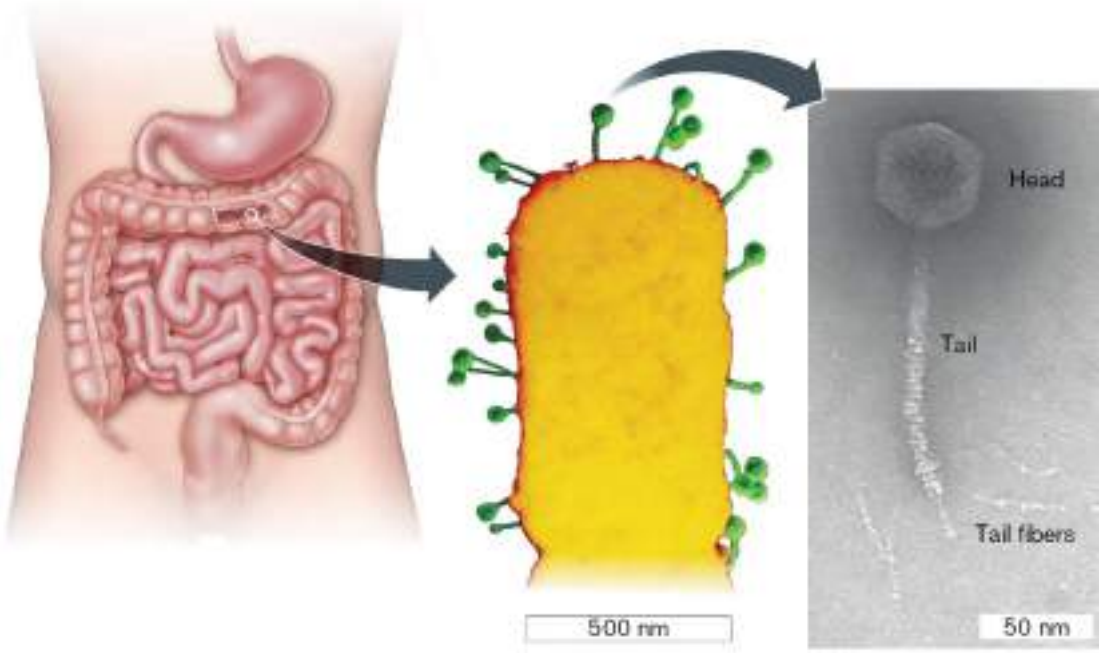


FIGURE 11.1 ■ Within the human intestine, *Escherichia coli* hosts phage lambda. Colorized SEM. **Inset:** Bacteriophage lambda, particle visualized by heavy-atom negative stain (TEM).

SCIENCE SOURCE
COURTESY OF ROBERT DUDA, UNIVERSITY OF PITTSBURGH

Historically, phage lambda's interaction with *E. coli* was the first living system simple enough to dissect at the molecular level. The "lambda switch" between lysis and lysogeny provided clues to the bacterial regulons presented in Chapter 10 and to molecular mechanisms of animals and plants. Today, the well-studied phage lambda provides tools for synthetic biosensors and even DNA chip devices.

Discovery of Phage Lambda

Bacteriophage lambda was discovered in 1950 by Esther Lederberg (1922–2006), pioneering bacterial geneticist at the University of Wisconsin–Madison and later Stanford University (**Fig. 11.2**). At the time of Lederberg's study, it was known that some kinds of bacteria

spontaneously lyse and release phages. But the mechanism, even the source of phage from within the cells, was a mystery.



FIGURE 11.2 ■ Esther Lederberg discovered phage lambda. Lederberg in her laboratory at Stanford, holding phage stocks in cotton-stoppered tubes.

THE ESTHER M. ZIMMER LEDERBERG TRUST. WWW.ESTHERLEDERBERG.COM

The strain *Escherichia coli* K-12 had been isolated in 1922 from a healthy patient's colon and was used for decades thereafter in teaching and research. In 1950, Lederberg was performing a genetic cross between two strains of *E. coli* K-12: a standard laboratory stock and a strain that she had mutagenized with ultraviolet light. After the two strains were mixed, the mutant strain formed colonies that were "nibbled and plaqued" (**Fig. 11.3A**). The plaques resulted from phage particles lysing the mutant strain. The mutant had lost its lambda prophage and was therefore susceptible to infection. The infecting phage particles came from the lysogenic *E. coli* K-12, which Lederberg had streaked across the susceptible mutant host.

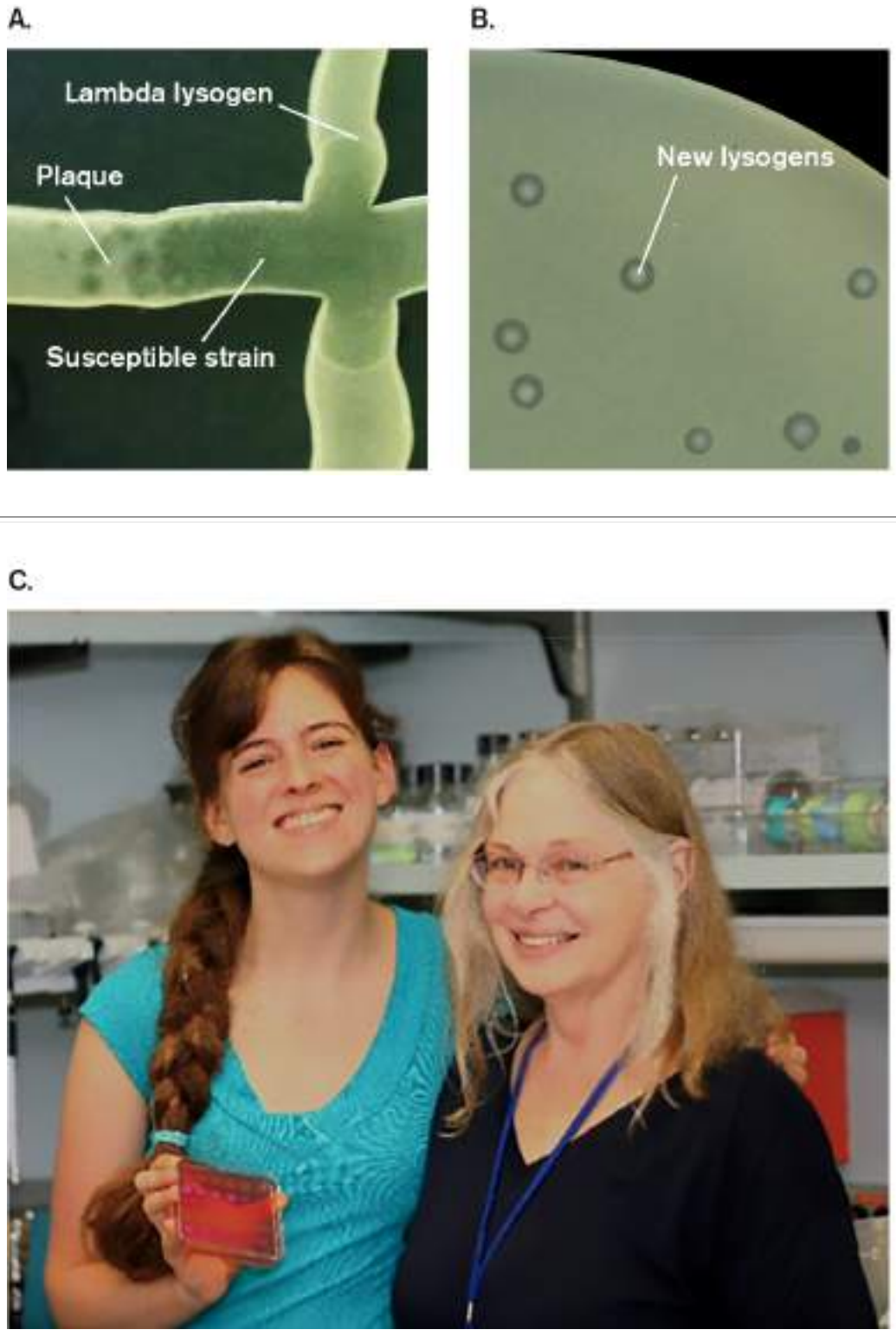


FIGURE 11.3 ■ Plate streak of a lysogen across a susceptible strain reveals phage lambda. A. The lysogenic *E.*

coli strain (vertical streak) releases phages that cause plaques in a susceptible *E. coli* strain (horizontal streak). **B.** Lambda plaques show a cloudy center where newly formed lysogens appear. **C.** Undergraduate Kathleen Morrill holds a plate of lambda *lacZ* reporter strains on MacConkey agar while standing next to Lynn Thomason at the National Cancer Institute.

COURTESY OF GÖKHAN TOLUN
COURTESY OF GÖKHAN TOLUN
COURTESY OF GÖKHAN TOLUN

Further experiments confirmed that at that time, all known stocks of K-12, the most widely used strain of *E. coli* , were lysogenic for phage lambda. Lysogeny means that the genome of a phage is incorporated into the genome of the host cell (discussed in Chapter 6). Phage particles then no longer exist as such, but the phage DNA replicates indefinitely within the host, as an integrated prophage . The presence of the prophage confers resistance to infection by the same type of phage. Phage lambda is thus called a "temperate phage," rather than a "virulent phage" such as T4, which always kills its host. But occasionally (about once in a million cells), a molecular signal tells the prophage genes to make progeny phages and lyse the cell. Thus, unknown to researchers at the time, a given tube of growing *E. coli* K-12 lysogen typically carried about a million lambda phages per milliliter. The phages released can form plaques on a strain that is not a lysogen (has no lambda prophage; **Fig. 11.3B**). The plaques have a cloudy center because some of the infected cells become lysogens, which start growing up where the infection began.

Lederberg and colleagues learned to "cure" the K-12 lysogen to eliminate the phage from stock cultures, including most strains used today. Other notable discoverers of the nature of lysogeny include French microbiologists André Lwoff (1902–1994) and François Jacob (1920–2013). The molecular basis of the lysis/lysogeny "switch" was elucidated by Mark Ptashne and colleagues (discussed next). Full regulation of the switch includes numerous viral and host proteins, whose details continue to be discovered by researchers such as Lynn

Thomason and her student Kathleen Morrill at the National Cancer Institute (**Fig. 11.3C**).

Thought Question

11.1 Plaques from phage lambda quickly fill with resistant lysogens. Could there be a different way for the host cells to become resistant to infection, without forming lysogens?

Phage Structure and Genome Replication

The phage lambda particle (virion) is typical of tailed phages of the siphophage family (**Fig. 11.4**). Various siphophages infect Gram-negative and Gram-positive bacteria. Siphophages commonly appear in soil and water, as well as in enteric communities.

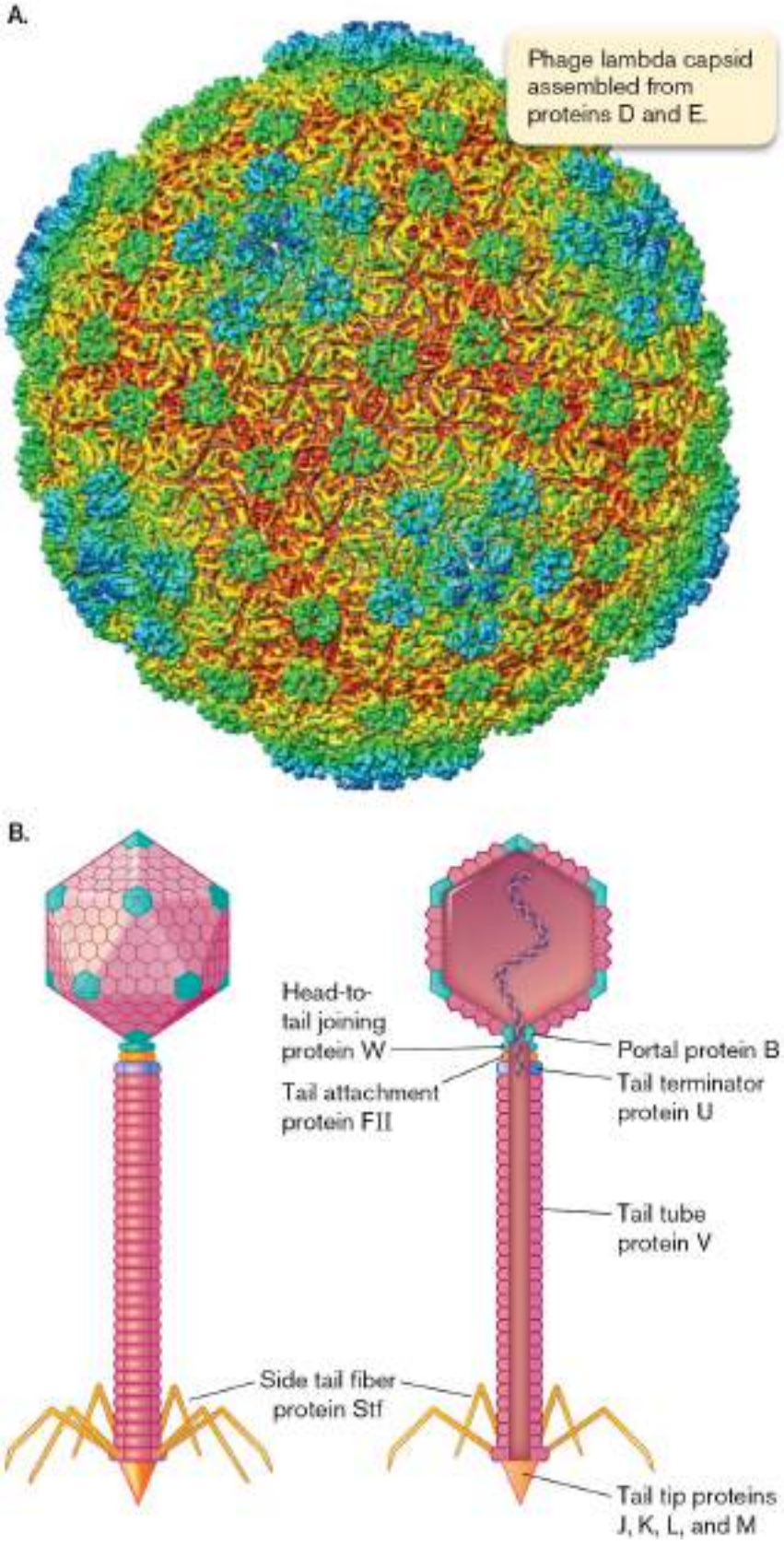

A.
Phage lambda capsid assembled from proteins D and E.
B.
Head-to-tail joining protein W
Tail attachment protein FII
Portal protein B
Tail terminator protein U
Tail tube protein V
Side tail fiber protein Stf
Tail tip proteins J, K, L, and M

FIGURE 11.4 ■ The lambda virion. A. Capsid complex, colored radially from center (red to blue). Cryo-EM. Green and cyan proteins lock capsid in place. **B.** Diagram of phage components, colored to match coding genes in the genome (see Fig. 11.5).

G. C. LANDER ET AL. 2008. *STRUCTURE* **16** :1399–1406
COURTESY OF ROBERT DUDA, UNIVERSITY OF PITTSBURGH

The head contains DNA. The part of a tailed phage that contains its genome is called the "head" or capsid. Usually the head consists of an icosahedral protein complex, equivalent to the capsid of a tail-less virion. The head of phage lambda contains a genome of double-stranded DNA. Thus, according to genome classification, phage lambda falls under Baltimore group I (see Chapter 6). Within the head or capsid, the DNA is spooled tightly. DNA is driven into the capsid by "inverse spooling"; that is, DNA winds around the capsid interior, laid down in compact parallel rings similar to those inside the herpes capsid (see the chapter-opening image). Then more DNA is driven in for several layers, until the central region is packed in a less ordered fashion.

The icosahedral protein coat is composed of several types of protein subunits, in regular repeating units that form 20 faces. **Figure 11.4A** shows a model of the phage lambda capsid that is based on cryo-electron tomography, a method that combines electron-microscopic images from frozen unstained virus particles (discussed in Chapter 2). The lambda model is colored to represent radial distance from the capsid center. The green and cyan proteins (proteins D and E) lock the capsid in place once it is filled with DNA. The capsid proteins provide remarkable strength, maintaining 60 atmospheres (atm) of pressure within the packed capsid. This pressure drives the eventual release of the DNA into the host cell. One vertex of the icosahedron is replaced by the tail connector. The tail connector comprises four proteins encoded by different genes (proteins B, FII, U, W). Each connector subunit is found in multiple copies that form a ring around the tail connector.

Tail tube. The tail itself consists of a long tube of 32 hexamer rings of subunit V. Remarkably, the exact length of the tail (the number of hexamers) depends on a "tape measure protein" (protein H). If we delete part of the gene that encodes the tape measure protein, the phage will assemble a shorter tail. On the other hand, if the gene is lengthened by addition of nucleotides, the phage will make a longer tail.

With respect to function, the tail of all siphophages is noncontractile (unlike the contractile tail of myophages such as phage T4). The tail tube has a tip and tail fibers, both of which help the phage attach to its host cell. Upon host attachment, the lambda DNA must uncoil and pass through the entire length of the tail to reach the host cytoplasm.

Genome. Phage lambda has a genome of 48.5 kilobase pairs (kb) that includes about 70 genes. In **Figure 11.5** , the left side of the genome encodes mainly components of the virion, including head, tail, and tail fibers. Most of these genes are color-coded to match their products, shown in **Figure 11.4B** . The right side of the genome includes enzymes for key functions such as host integration, DNA replication and lysis, and the famous lysis/lysogeny switch (discussed shortly).

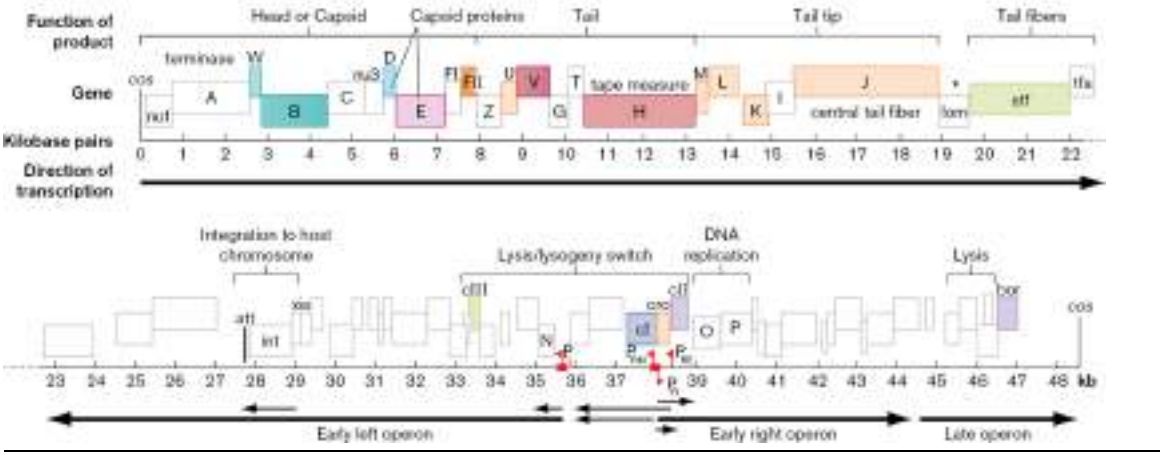


FIGURE 11.5 ■ Genome of phage lambda. For each gene, vertical offset indicates the reading frame. P_L = leftward

promoter; P_R = rightward promoter; P_{RE} = repressor expression promoter; P_{RM} = repressor maintenance promoter.

Within a phage head, the packaged genome is linear. The left and right ends each possess *cos* , a sequence of 200 bp that was cleaved by a terminase enzyme when the phage DNA was packaged. The cleavage generated a staggered cut, leaving short stretches of bases unpaired (**Fig. 11.6**). As the phage DNA now enters the new host cell, the *cos* unpaired ends anneal together, and their backbones are sealed by DNA ligase (a replication enzyme discussed in Chapter 7). The genome is now circular, as the sealed ends form a long operon transcribed "rightward" from P_R , all the way through the virion structural components (**Fig. 11.5**). Another operon is transcribed "leftward," from P_L . Note that both the P_R and P_L operons use all three reading frames, and that some coding genes overlap. Long messenger RNA (mRNA) transcripts with gene overlap in three reading frames are a common feature of viral genomes.

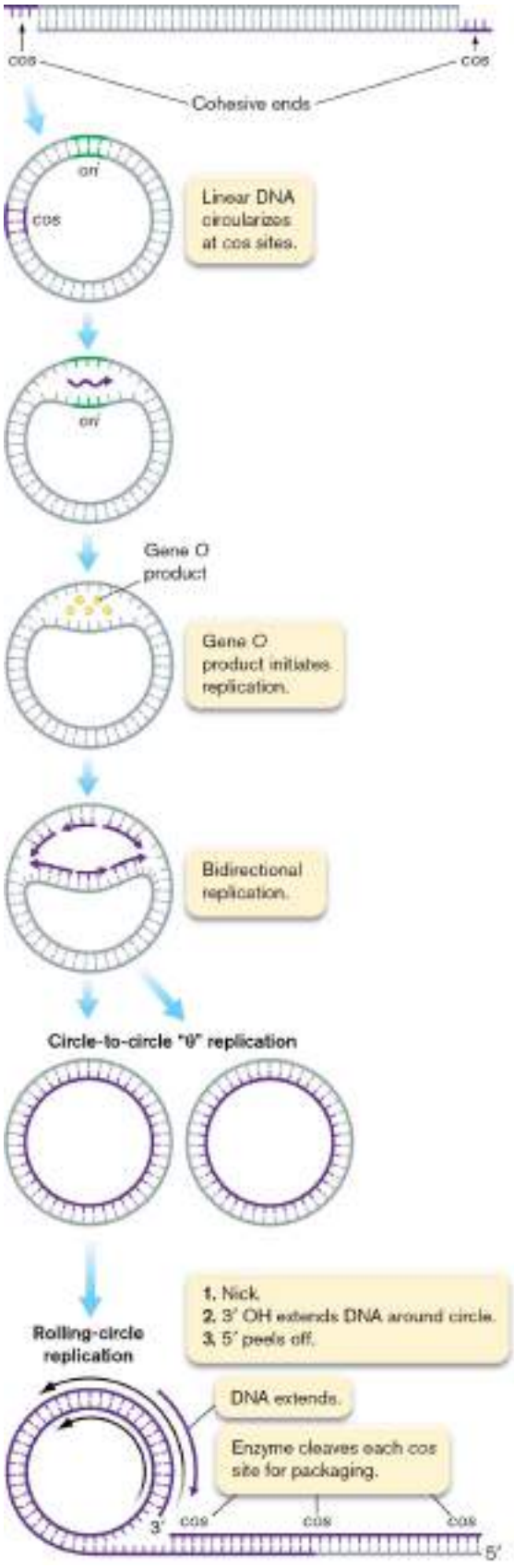

cos
cos
Cohesive ends
ori
cos
Linear DNA circularizes at cos sites.
ori
Gene O product
Gene O product initiates replication.
Bidirectional replication.
Circle-to-circle "θ" replication
Rolling-circle replication
1. Nick.
2. 3′ OH extends DNA around circle.
3. 5′ peels off.
DNA extends.
Enzyme cleaves each cos site for packaging.
3′
cos
cos
cos
5′

FIGURE 11.6 ■ Replication of phage lambda. Initially, the linear genome circularizes and replicates bidirectionally. Then it replicates multiple genomes end to tail, forming a concatemer.

Besides the P_R and P_L promoters, which function during the lytic cycle, other promoters are used for key events of the lysis/lysogeny switch.

During the lytic cycle, the circularized DNA molecule undergoes several rounds of bidirectional replication via host DNA polymerase (**Fig. 11.6**). The completed circles are then nicked (cleaved on one strand), generating a 3′ OH end. The 3′ OH end of DNA serves as a primer for rolling-circle replication . Rolling-circle replication is commonly used by plasmids (see Chapter 7), as well as by circularized viral genomes such as that of herpesviruses (see Section 11.5).

In the rolling-circle process, one strand of DNA extends its 3′ end continually around the circular template, while the 5′ end "rolls away." The 5′ extension generates a long line of tandemly repeated genomes called a concatemer . The complementary strand of DNA fills in later. Once the long, double-stranded molecule is complete, the terminase enzyme cleaves the concatemer into pieces that fit into phage head coats. For phage lambda, the genomes are cleaved at the *cos* site, which acts as a signal to package the progeny genome into the protein head coat. Cleavage restores the two *cos* ends to form linear genomes coiled into the phage head.

Thought Question

11.2 What advantages does rolling-circle replication offer a phage, compared with bidirectional replication?

Phage Attachment and Infection

To infect a host cell, a virus needs to attach to the host surface and insert its genome into the cytoplasm (discussed in Chapter 6). Most types of bacteriophages extrude their genome from the head or capsid and thread it across the bacterial envelope while leaving the capsid outside. An exception is phage M13, whose filamentous capsid penetrates the entire envelope to replicate slowly within the cytoplasm. Phage lambda, however, uses the more common mechanism of binding to a specific cell-surface receptor. The phage adsorbs (attaches) to the receptor by contact with its tail fibers.

The receptor for phage lambda is the *E. coli* uptake complex for maltose (glucose dimer) and other short-chain sugars. In the colon, *E. coli* obtains these short glucose chains from anaerobes such as *Bacteroides* species that break down large, complex polysaccharides from plant material (discussed in Chapters 13 and 21). Sugar chain uptake is crucial for *E. coli* , so the bacteria are unlikely to lose the uptake complex by evolution, even though phage lambda takes advantage of it for infection.

Phage lambda binds specifically to maltose porin, the outer membrane pore that transports maltose into the cell (**Fig. 11.7**). Porins are a large family of proteins that share a distinctive "beta barrel" structure (presented in Chapter 3). The beta barrel of maltose porin (blue in **Fig. 11.7**) is buried in the outer membrane. The phage-binding sites (green) were identified by amino acid substitution mutations that prevent phage binding and confer host resistance to lambda. The maltose porin was first called "lambda receptor," LamB, because it was discovered by its function as the lambda receptor. Of course, the protein actually evolved in the host as a way to obtain nutrients.

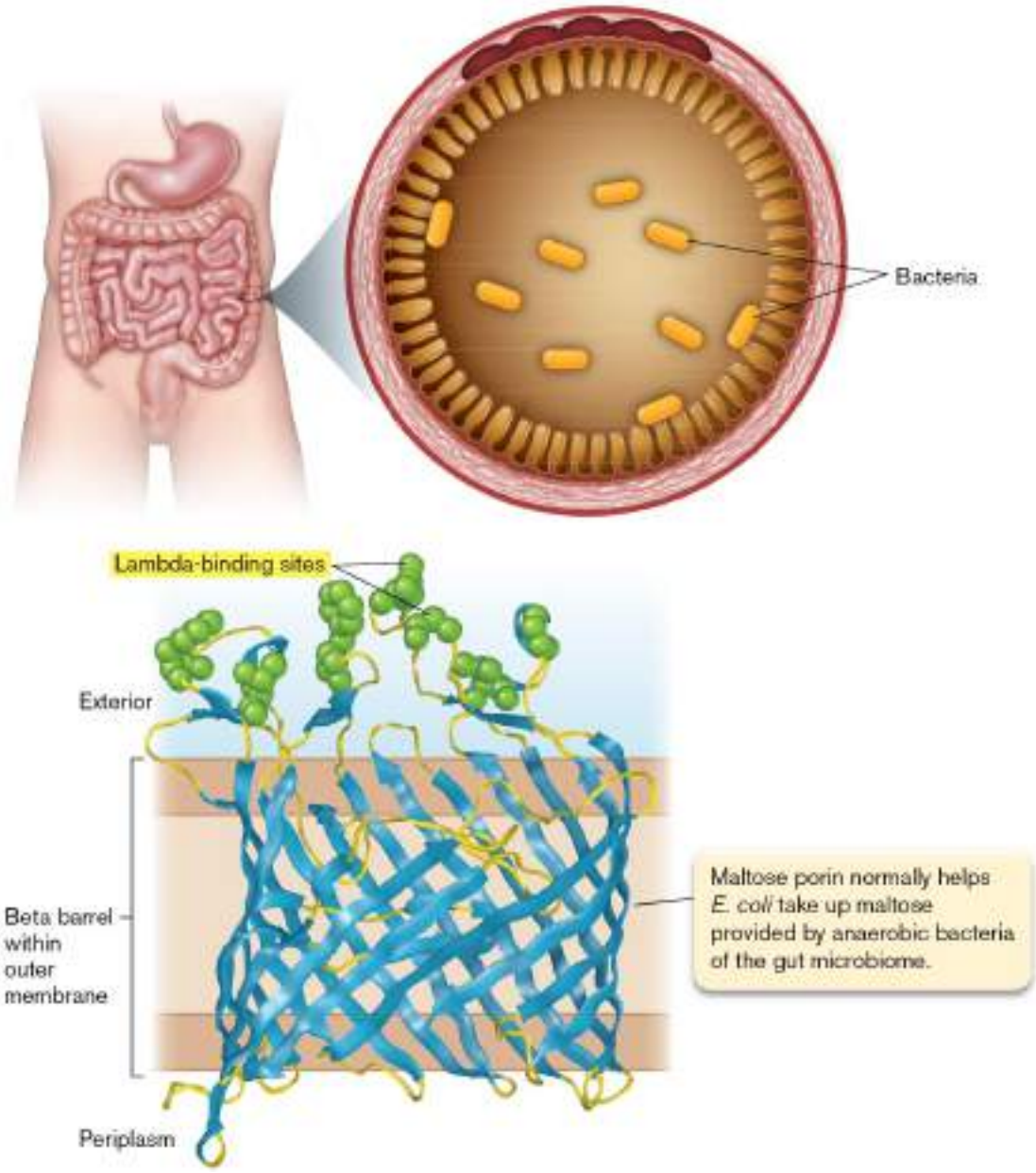


FIGURE 11.7 ■ Host receptor: maltose porin. *E. coli* bacteria in the human gut use maltose porin to obtain maltose. The maltose porin is embedded in the outer membrane of *E. coli* . (PDB code: 1MAL)

Attaching to the cell surface is just the first challenge that the phage faces to establish infection (**Fig. 11.8** , step 1). How does the phage get its DNA all the way across the periplasmic space and the inner membrane? Unlike phage T4 (whose tail contracts to expel DNA

under pressure), phage lambda uses an extrusion mechanism that is not fully understood. Somehow the DNA threads across the periplasm and through the maltose transport complex of the inner membrane to reach the cytoplasm (step 2). Within the cytoplasm, the staggered ends of the *cos* sites anneal and are ligated, circularizing the genome (step 3). Now, the host RNA polymerase can begin to transcribe the phage operons.

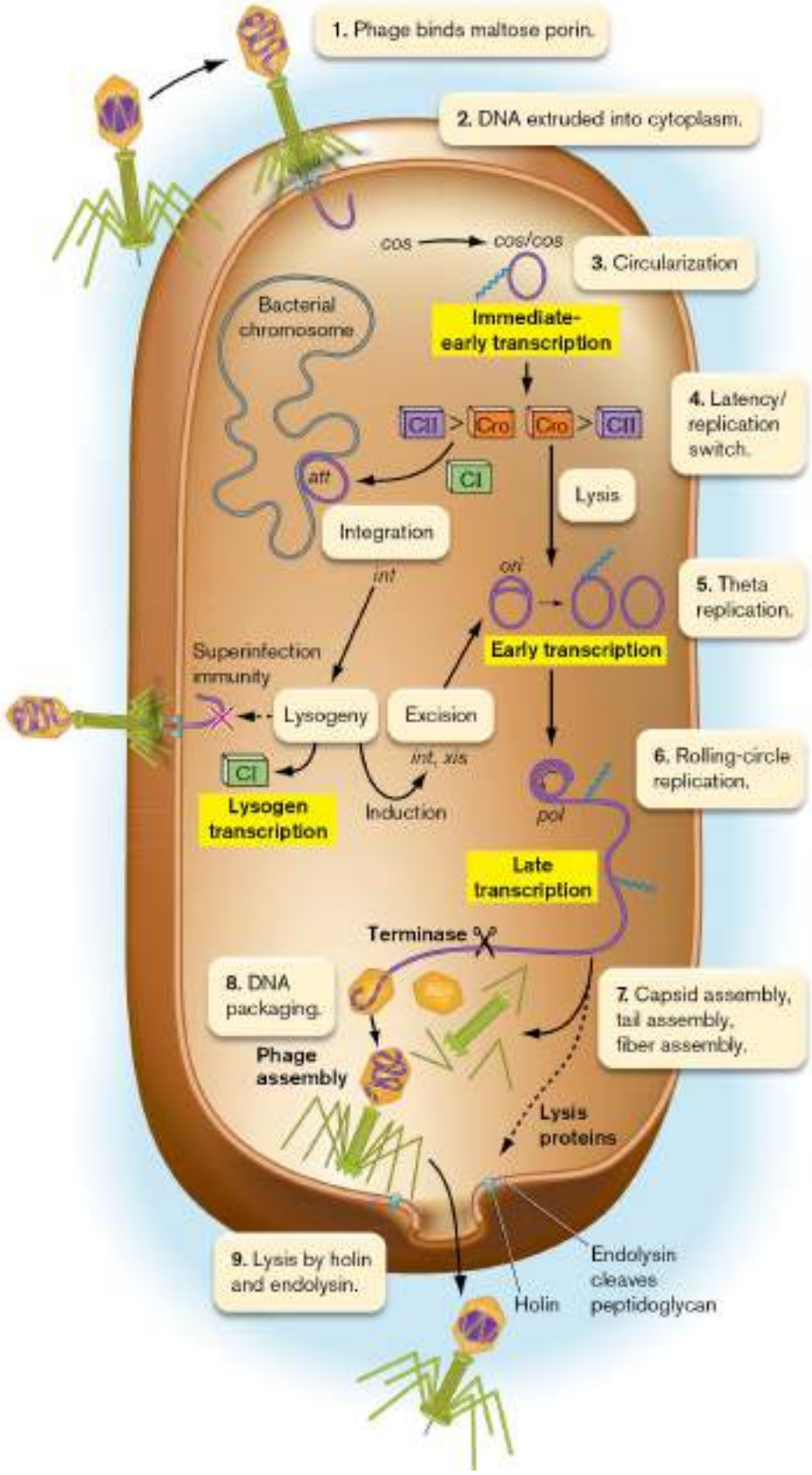

1. Phage binds maltose porin.
2. DNA extruded into cytoplasm.
cos
cos/cos
3. Circularization
Bacterial chromosome
Immediate-early transcription
CII
Cro
Cro
CII
4. Latency/ replication switch.
att
CI
Lysis
Integration
int
ori
5. Theta replication.
Early transcription
Superinfection immunity
Lysogeny
Excision
int, xis
CI
Lysogen transcription
Induction
pol
6. Rolling-circle replication.
Late transcription
Terminase
8. DNA packaging.
7. Capsid assembly, tail assembly, fiber assembly.
Phage assembly
Lysis proteins
9. Lysis by holin and endolysin.
Endolysin cleaves peptidoglycan
Holin

FIGURE 11.8 ■ Phage lambda infection of *Escherichia coli* K-12.

At first, though, the newly introduced phage genome supports transcription of just a few key proteins, most notably the "control proteins" Cro and CII (**Fig. 11.8** , step 4):

- **Cro protein leads to lysis.** Small amounts of Cro activate the lytic cycle.
- **CII protein leads to lysogeny.** Small amounts of CII block expression of Cro and lytic proteins, and CII induces expression of CI (known as "lambda repressor").

So, which protein wins—Cro or CII? The answer depends on numerous factors in the gut environment, whose signals combine to lead to one decision or the other. One such signal is a surge in nutrients, such as when you eat a meal, sending rich organic substrates down to your intestines. High nutrient concentrations cause the *E. coli* protease HflB to cleave the phage protein CII, leaving Cro to induce the lytic cycle. Low nutrient concentrations inhibit HflB, leaving CII available to block Cro.

Alternatively, suppose multiple phage particles coinfect the cell at the same time. This event implies that the host population is outnumbered by phages, and thus progeny phages will have poor opportunities to find a host.



For this reason the phage has evolved a mechanism to avoid lytic reproduction: The multiple phages express a higher level of CII, some of which now evades the host protease and blocks Cro. When CII blocks Cro, the fortunate host cell becomes a lysogen instead of lysing. A phage genome integrates in the host genome as a prophage, at the *att* site. The phage DNA integrates with *att* by site-specific recombination; that is, recombination between the backbones of two DNAs that share a short sequence in common (discussed in Chapter 9).

The integrated prophage now expresses only a few proteins, including CI (lambda repressor). CI repressor prevents the lytic cycle by blocking transcription of lytic promoters. The CI repressor also

prevents superinfection by other lambda phages. This is what happens to most of your intestinal bacteria between meals: The bacteria stay in stationary phase with their lysogenic phages repressed.

The phage lambda CI repressor is famous because its binding to DNA was the first protein-DNA binding event to be described (**Fig. 11.9**). Mark Ptashne at Memorial Sloan Kettering Cancer Center (**Fig. 11.9A**) showed how CI protein forms a dimer that binds a specific DNA sequence. Each CI subunit binds DNA through an interaction between an alpha helix of the protein and a major-groove sequence of DNA (**Fig. 11.9B**). Similar alpha helix binding to the major groove mediates the function of many genetic regulators of animals and plants, as well as bacteria.

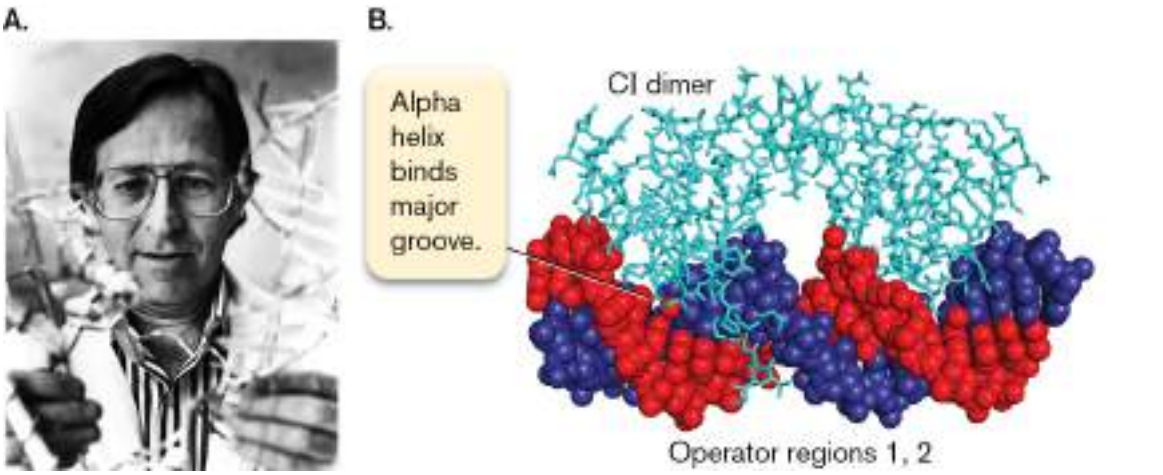


FIGURE 11.9 ■ CI repressor binds DNA. A. Mark Ptashne worked out the binding of CI repressor to DNA and the mechanism of the lysis/lysogeny switch. **B.** The CI repressor dimer binds the operator sequence by fitting the DNA sequence at the major groove. (PDB code: 1LMB)

NIH

When phage lambda infects *E. coli* , what if the Cro control wins, instead of CI? Cro protein represses CI expression, thus promoting the lytic cycle. The lytic cycle leads to lysis (host cell destruction, and release of progeny phages). The circularized phage DNA replicates,

first bidirectionally (**Fig. 11.8** , step 5) and then by rolling-circle replication (step 6). Structural proteins are synthesized and assembled to form the empty capsid (or head coat), tail tube, and tail fibers (step 7). The *cos* end of the DNA concatemer gets stuffed into a progeny head coat (step 8). The stuffed portion of the concatemer is then cleaved at the next *cos* site by a phage enzyme called terminase. Terminase also helps attach the filled head to the tail tube. When most of the progeny phages have assembled, a phage-encoded protein called holin punctures the inner cell membrane, providing a channel through which endolysin reaches the cytoplasm. Endolysin cleaves peptidoglycan (step 9, lysis). Now, holes open across the envelope, and phage particles emerge from the destroyed cell.

Thought Questions

11.3 A researcher adds phage lambda to an *E. coli* population whose cells fail to express maltose porin. After several days, the *E. coli* are now lysed by phage. What could be the explanation?

Lysogeny: To Lyse or Not?

As a lysogen, the host *E. coli* replicates just as it did without any prophage integrated. The prophage expresses CI repressor, which blocks nearly all expression of phage genes dangerous to the cell and prevents superinfection by other phage lambda particles. Yet at any time, a spontaneous event may override the CI repressor and trigger a lytic cycle. The lysis/lysogeny decision centers around CI and Cro proteins, with multiple other regulators that modify their function, only some of which are shown in **Figure 11.10** . Such complexity is evidence of a lengthy evolution, in which multiple "adjustments" had time to occur in both phage and host genomes. The result is a control

network highly responsive to diverse signal inputs from the environment.

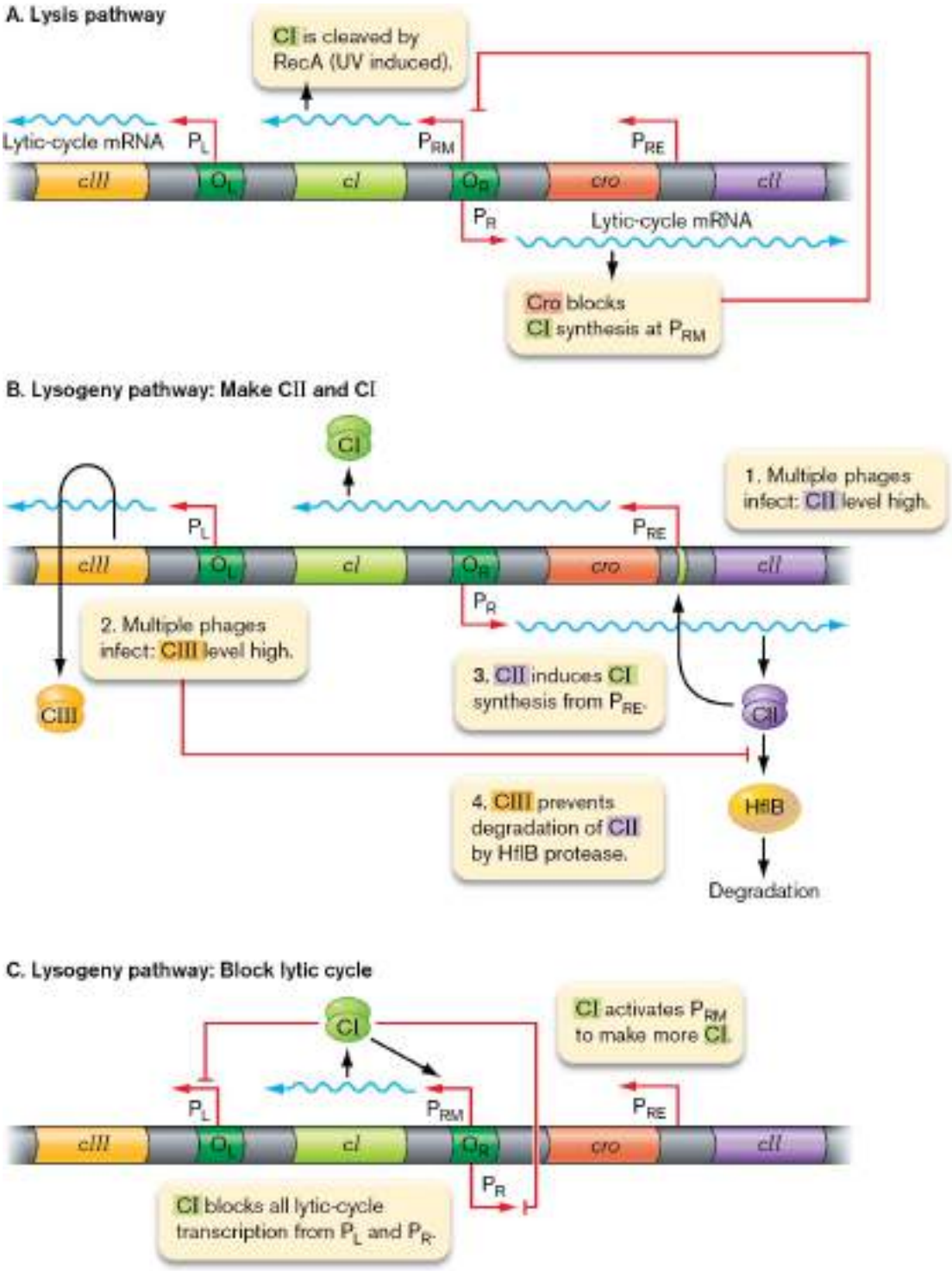


FIGURE 11.10 ■ Lysogeny: to lyse or not? A. The CI repressor maintaining lysogeny is cleaved by RecA (induced by UV exposure). Cro blocks CI synthesis, allowing expression of the lytic promoters P_L and P_R. **B.** If multiple phages infect simultaneously, the genome expresses enough CII protein to bind P_{RE}, activating expression of CI. **C.** The CI protein blocks promoters P_L and P_R (lytic cycle) and activates P_{RM} (to make more CI).

Induction of lytic cycle. During lysogeny, the CI repressor blocks expression of nearly all other genes from the promoters P_L and P_R (**Fig. 11.10C**). These two promoters, particularly P_R, express most of the phage structural proteins and lytic enzymes from the circularized lambda genome (**Fig. 11.5**).

But the number of repressor molecules present in a cell is small enough for fluctuation to lead to rare events. Occasionally, perhaps one in 10^7 cells, the CI concentration is lowered to a level at which P_R becomes exposed, enabling expression of Cro—which activates lysis. CI is further decreased by host RecA, which cleaves this protein along with many host repressors (**Fig. 11.10A**). RecA is activated by DNA damage during UV light exposure, so UV light can induce lysis in 100% of a lysogenic population.

When the Cro protein (lysis activator) is expressed, it binds to both operators (O_L and O_R) for the P_L and P_R operons, respectively. Cro up-regulates these lysis operons during the lytic cycle and, at the same time, blocks expression of CI from promoter P_{RM}. Thus, no further repressor can be made to block phage production, and the cell is committed to lysis.

Maintaining lysogeny. To avoid lysis requires continual expression of CI repressor (**Fig. 11.10B**). Early in lysogeny, CI is expressed from promoter P_{RE}. CI expression requires P_{RE} to bind CII protein. But CII protein is vulnerable to cleavage by host HflB (in a high-

nutrient medium). HflB cleavage can be prevented by CIII, a protein expressed early from promoter P_L.

Once CI protein attains sufficient concentration, it activates its own expression from a different promoter, P_{RM}. P_{RM} activation requires the binding of CI dimers to O_R. Besides activating P_{RM}, the CI binding also blocks expression of P_R, the major lytic operon, as well as the leftward-transcribing operon from P_L. From then on, CI maintains its own expression while preventing induction of lysis. Lysis occurs only when stress such as DNA damage activates RecA to cleave CI.

The full repression of lysis requires eight molecules of CI in all: a pair of dimers at O_R, and another pair of dimers at O_L (**Fig. 11.10C**). The two pairs of dimers actually bind each other as an octamer, with DNA looped between them (not shown). The dimers work together so that the binding of two regulators is stronger than the sum of the individual regulators binding DNA. This cooperative binding increases the on/off character of a molecular switch, lessening the occurrence of partial states in between. Another example of cooperative molecular regulation is that of the LacI repressor in Chapter 10 (see **Fig. 10.7**).

From Mobile Genes to Synthetic Biology

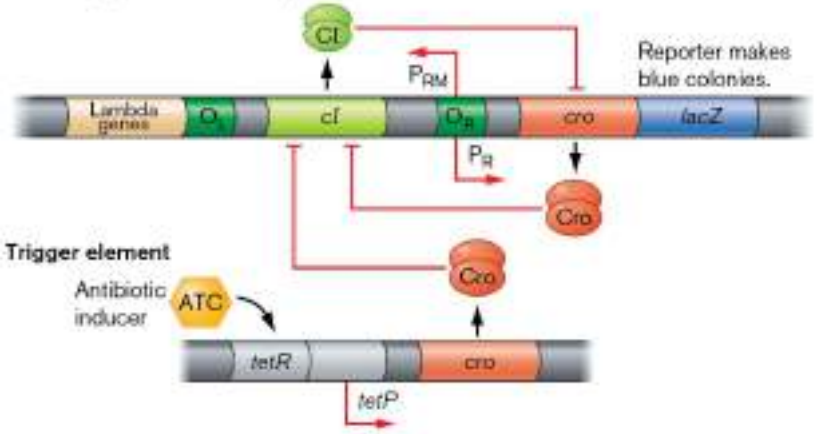
Within natural microbiomes such as that of our intestine, lysogenic phages carry genes that offer useful functions to their host bacterium. An example is the *bor* gene (found at the right end of the lambda genome; **Fig. 11.5**). The *bor* gene encodes Bor lipoprotein, which resides in the outer membrane of the host bacterium. The presence of Bor protects a lysogen from destruction by the serum complement cascade (a form of innate immunity, described in Chapter 23). The location of the *bor* gene near the end of the phage genome suggests that it was picked up from an ancestral host by a process of specialized transduction (see Chapter 9). In specialized transduction, a prophage initiates lysis by copying its genome out of the integration site, and it picks up a small adjacent piece of host DNA. The host

DNA is then copied into progeny phage and is transferred to the next host infected.

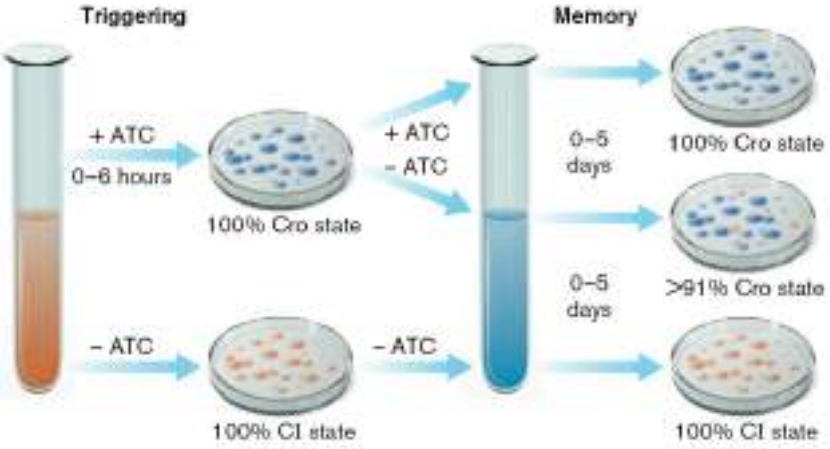
Today, our genetic engineering can make artificial use of phages to carry genes with properties useful to us. We use genetic elements of the lambda switch for synthetic biology; in effect, we imitate natural gene transfer mechanisms to construct bacteria with functions useful to us. For example, Pamela Silver's students at Harvard University used a CI/Cro switch to build a bacterial recorder that detects an environmental signal within the intestine. This kind of bacterial device could be developed as a detector of signals in the human gut, such as cancer molecules.

Silver's bacterial recorder possesses a stripped-down version of the CI/Cro switch (**Fig. 11.11A**). The *cro* gene is fused to *lacZ* , whose product generates blue colonies when expressed on indicator plates. (Gene fusion is presented in Chapter 2 ; see **Fig. 2.29** .) But *cro* expression from promoter P_R is blocked by CI repressor, which in turn is expressed from a different promoter, P_{RM} . Without *cro-lacZ* expression, there is no blue color—so the net result is colonies that are white.

A. Memory element with reporter



B. Antibiotic triggers Cro; makes colonies blue



C. Within a mouse, bacteria record antibiotic exposure

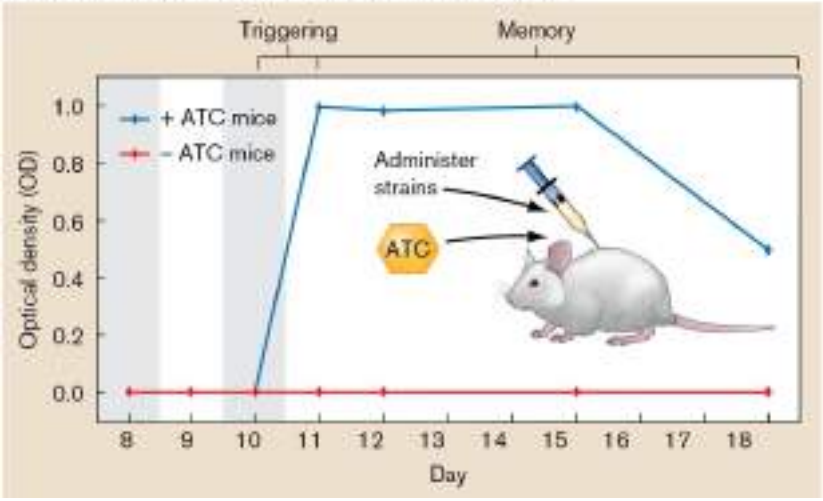


FIGURE 11.11 ■ Antibiotic reporter bacterium uses a CI /Cro switch. A. The CI /Cro switch was engineered to record exposure to an antibiotic (ATC). The Cro "memory" is triggered by an inducer that turns off CI expression over many cell generations. **B.** The ATC signal turns colonies blue. Blue colonies persist for 5 days. **C.** The bacterial detector works when administered to mice. Seven days after addition and removal of ATC, the engineered bacteria still make blue colonies (Cro state).

Source: Jonathan Kotula et al. 2014. *PNAS* **111** :4838.

The same bacteria contain another fragment of the lambda switch, the "trigger element," in which Cro protein is expressed from a gene under control of the antibiotic-inducible promoter *tetP* . Now, suppose we add the inducing antibiotic (ATC) to the bacteria. The *tetP-cro* trigger element now expresses Cro, which represses expression of CI protein by binding P_{RM} . With CI expression turned off, the *cro-lacZ* fusion indefinitely expresses its protein, causing blue colonies (**Fig. 11.11B**). The blue phenotype recurs for many generations.

What happens when we put this bacterial recorder into the microbiome of a mouse? First, to colonize the mouse intestine, the bacterium needs to include a selective gene for competitive advantage within the microbial community—a gene conferring resistance to streptomycin. The mice are treated with streptomycin until the bacterial recorders are established (**Fig. 11.11C**). Then the researcher adds the test antibiotic ATC. After ATC is removed, bacteria are sampled from mice fecal pellets. For up to 7 days after ATC removal, blue colonies appear, showing that the bacteria record ATC exposure and report it long after the signal is gone. Further applications of synthetic biology are described in Chapter 12 .

We now turn to viruses that infect humans, focusing on three whose biology is well studied: influenza virus, a negative-strand RNA virus; human immunodeficiency virus (HIV), the retrovirus that causes AIDS; and herpes simplex, a double-stranded DNA virus

causing oral and genital herpes. Viruses of humans show mechanisms that resemble those of bacteriophages. For example, retroviruses integrate the DNA copy of their genomes into the host cell genome, analogous to lysogeny by phage lambda. Other viruses, such as influenza, have no known latent form. The influenza virus is highly virulent and lyses infected cells with little or no latent state.

To Summarize

- **Bacteriophage lambda was discovered in an *E. coli* lysogen from a human colon.** An *E. coli* lysogen released phage that infected a sensitive strain. The human gut microbiome is full of phage lysogens.
- **The phage lambda virion** consists of a head containing its DNA genome and accessory proteins, a tail composed of an internal tube, and tail fibers.
- **Phage lambda binds to the host maltose porin.** Phage DNA is inserted into the cytoplasm, where early phage genes are expressed.
- **Control proteins bind to DNA, leading to lysis or lysogeny.** Single-phage infection (Cro protein) leads to lysis, whereas multiple-phage infection (CII protein) more likely leads to lysogeny.
- **In a lytic cycle, rolling-circle replication generates progeny genomes.** The progeny genomes are packaged into head coats and then cleaved from the concatemer; the filled heads are attached to tails. Late-expressed proteins lead to lysis.
- **CII up-regulates expression of CI repressor, which maintains lysogeny.** CI then induces its own expression, while repressing expression of Cro and proteins of the lytic cycle. By this kind of regulation, many human gut bacteria switch between lysis and lysogeny.

- **The CI/Cro switch is used for synthetic biology.** The molecular switch is built into biomedical devices.

Glossary

lysogeny
: A viral life cycle in which the viral genome integrates into and replicates with the host genome but retains the ability to initiate host cell lysis.

prophage
: A phage genome integrated into a host genome.

rolling-circle replication
: A form of DNA replication that proceeds in one direction around a circular template, making tandem copies in a linear array (concatemer). The copies are later cleaved and circularized.

concatemer
: A long line of tandemly repeated genomes; commonly formed during rolling-circle replication.

lysis
: Also called *burst* . The rupture of the cell by a break in the cell wall and membrane.

Fig. 10.7

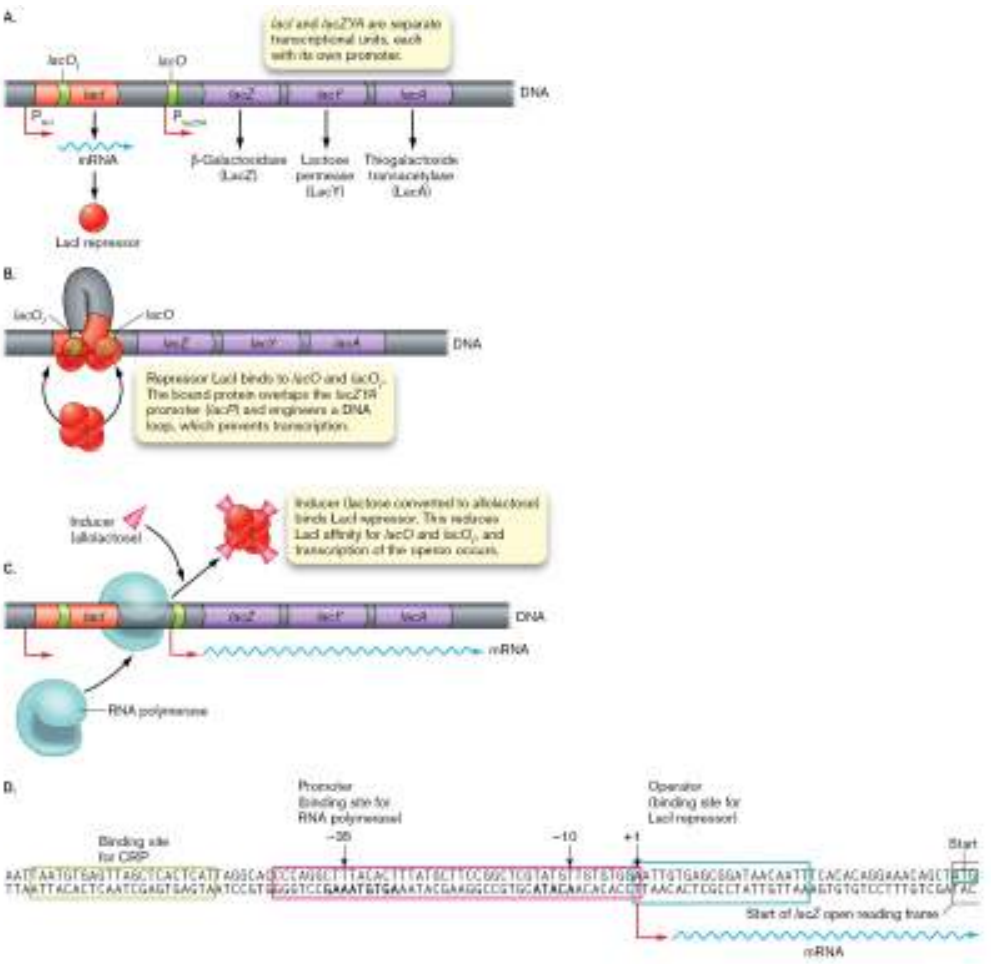


FIGURE 10.7 ■ Transcriptional induction of the lactose operon. A. Organization of the operon. Bent arrows mark promoters. Green indicates LacI protein-binding sites on DNA. **B.** The LacI tetrameric repressor binds to specific DNA sites (the operators *lacO* and *lacO* $_I$). **C.** The inducer allolactose (an altered form of lactose made by low levels of beta-galactosidase) removes the repressor LacI and allows expression of *lacZYA* . **D.** DNA sequence of the *lac* control region.

Fig. 2.29

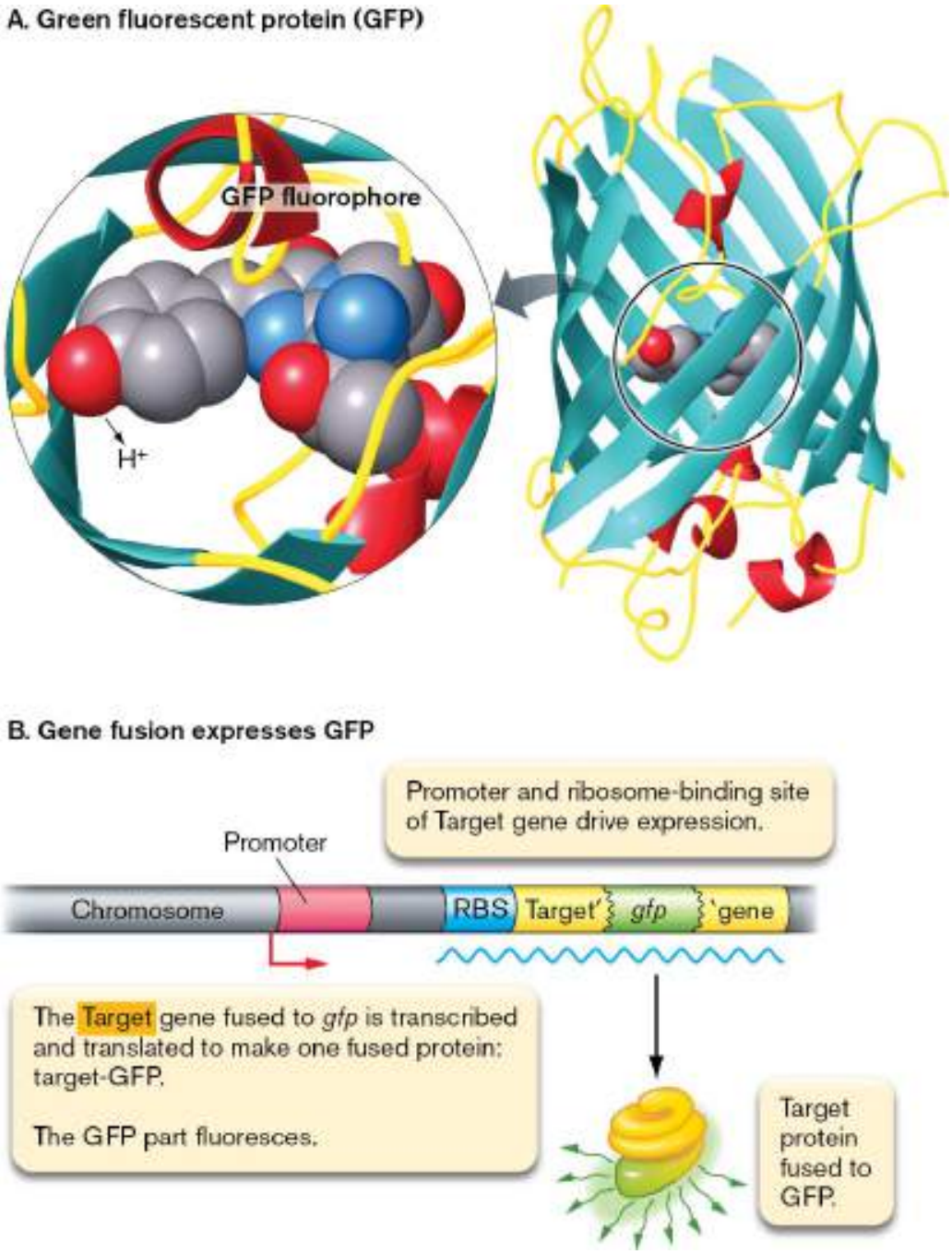


FIGURE 2.29 ■ The fluorophore green fluorescent protein (GFP). A. Green fluorescent protein (GFP) is expressed endogenously by the cell. **Blowup:** Three GFP amino acid residues (serine, tyrosine, and glycine) condense to form the fluorophore. **B.** The gene encoding GFP can be fused to a target gene (Target'- *gfp*). The fused gene then expresses a fused protein in which the GFP portion fluoresces.

The fluorescent protein is expressed under control of the target gene promoter and ribosome-binding site (RBS).

11.2 Influenza Virus: (–) strand RNA Virus

Influenza virus is a major human pathogen, causing up to half a million deaths per year worldwide, as well as pandemics such as the H1N1 "swine flu" in 2009. Research focuses on the mechanisms of influenza infection as targets for new antiviral agents. But human viruses are challenging to study—more so than bacteriophages. Viruses infecting human cells require complex replication cycles among the compartments of a eukaryotic cell. Unlike bacteriophages, the progeny virions of a human virus must navigate our organ systems to maintain infection and achieve transmission to new hosts. For example, avian influenza strains poorly transmit to humans because the cell-surface receptors these strains need are hidden deep within the human respiratory tract. Human and other animal viruses face the defenses of our immune system (see Chapters 23 and 24). Even for the deadliest strains, a large proportion of infected people eliminate the virus without symptoms (asymptomatic infection). Why do some people fight off the virus, whereas others who appear equally healthy succumb? Many mysteries remain to challenge the research virologist.

History of Influenza

Besides the seasonal strains, influenza shows a cyclic appearance of strains that cause pandemic mortality. The famous pandemic of 1918, for example, infected 20% of the world's population and killed more people than those who died in World War I. The 1918 strain arose as a mutant form of an influenza strain infecting birds. Another strain infecting humans, H3N2, emerged from avian influenza strains in 1968, causing high mortality in humans. We discuss shortly the distinctive basis of influenza "reassortment" mutations and how such sudden, virulent strains emerge.

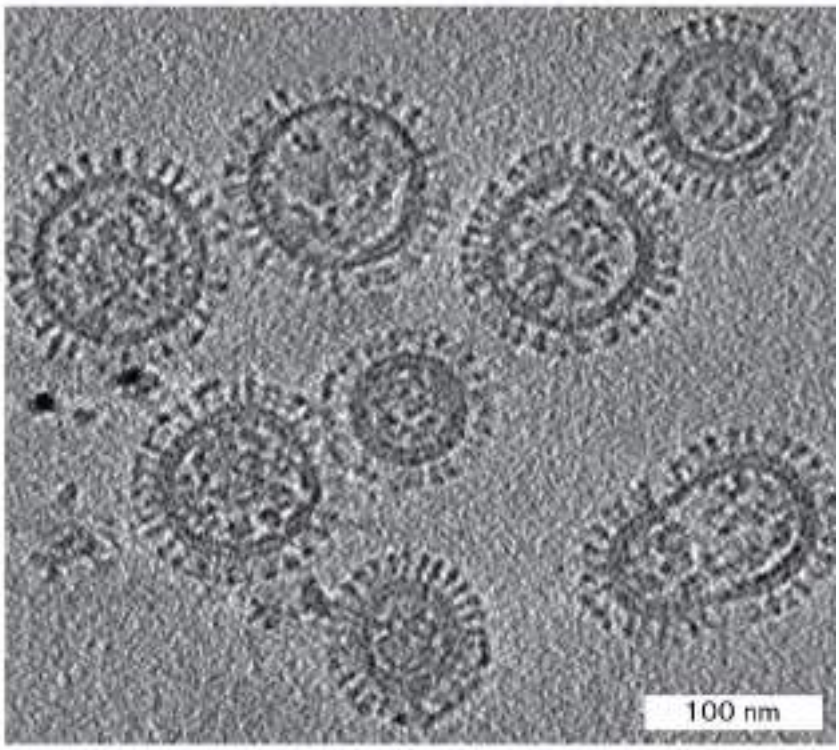
In 2009, a highly transmissible strain emerged that combined genes of human-infecting strains with those of strains infecting swine. This H1N1 variant spread rapidly around the world, infecting many young people including college populations. Fortunately, this strain caused relatively mild illness. A future strain, however, might emerge combining the high transmission seen in swine flu with the high human mortality seen in the avian strain. Modern human travel patterns accelerate the spread and mixing of animal and human influenza strains.



Virion Structure and Genome

The influenza virion has an asymmetrical structure (**Fig. 11.12**). The structure of influenza A virus (IAV) was imaged by Audray Harris and colleagues at the National Institutes of Health through use of cryo-electron microscopy (cryo-EM). The envelope appears as a hollow sphere, sometimes elongated, studded by peg-shaped protein complexes of hemagglutinin (HA) trimer and neuraminidase (NA) tetramer. These envelope proteins are found in numbered variants, giving rise to the infamous acronyms of certain influenza strains such as the H1N1 pandemic strains of 1918 and 2009. The shape of HA and NA at higher resolution within the envelope was determined by cryo-electron tomography (discussed in Chapter 2). Harris used this technique to devise a 3D model of the virion (**Fig. 11.12B**). Within the envelope surface, the HA and NA proteins collect in uneven "domains" lacking symmetry, even with empty patches of envelope devoid of proteins. This asymmetry may increase influenza's variability and generation of new strains.

A.



B.

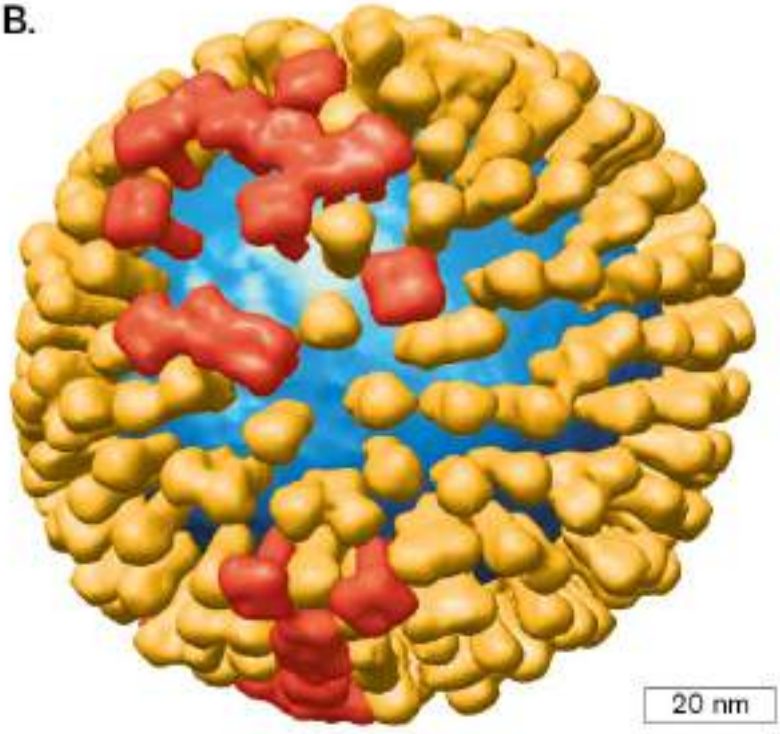


FIGURE 11.12 ■ Influenza A virus (IAV) virions imaged by cryo-EM. A. IAV particles imaged by cryo-EM show varying size, shape, and distribution of envelope proteins. **B.** Virion model showing HA tetramers (red) and NA trimers (light orange).

A. HARRIS ET AL. 2006. *PROC NATL ACAD SCI USA* **12** :19123–7

A. HARRIS ET AL. 2006. *PROC NATL ACAD SCI USA* **12** :19123–7

The IAV virion has no icosahedral capsid. Instead, the (–) strand RNA is packaged by winding around nucleocapsid proteins (NPs) (**Fig. 11.13A**). The term "nucleocapsid" refers generally to proteins coating a viral genome and packaged within or as part of the virion.

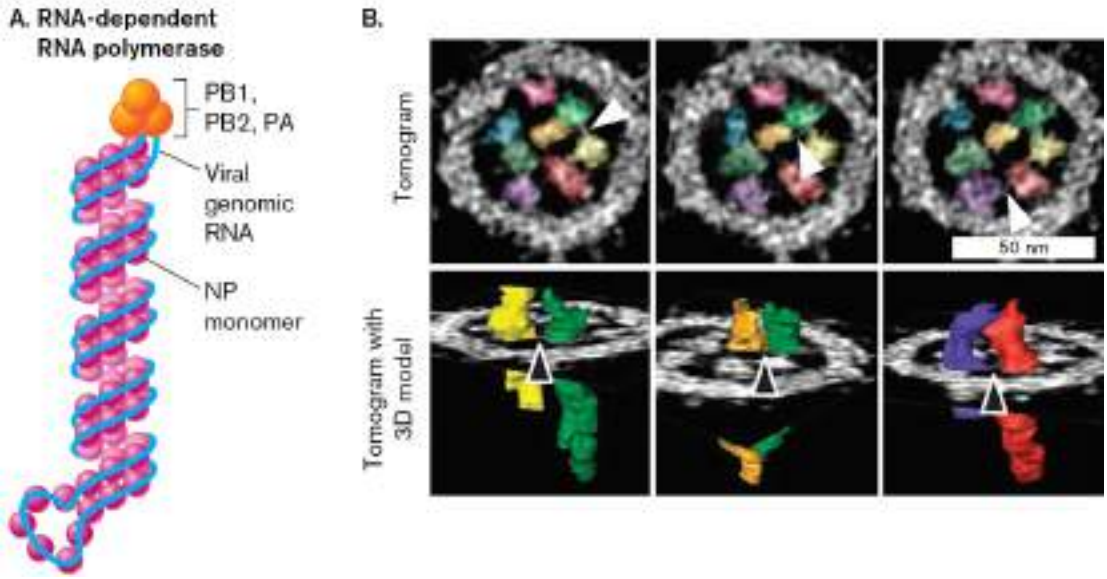


FIGURE 11.13 ■ Influenza ribonucleoprotein (RNP) complexes are linked and packed in the virion. A. An RNP complex includes the RNA chromosome segment helically wrapped around NP monomers and attached to an RNA-dependent RNA polymerase (PB1, PB2, PA). **B.** Influenza RNP complexes packed within a virion (colorized cryo-EM). Three tomography sections reveal links between specific RNPs (arrowheads).

Source: Part A modified from Amie Eisfeld et al. 2015. *Nat. Rev. Microbiol.* **13** :28.

TAKESHI NODA ET AL. 2012. *NAT. COMMUN.* **3** :639

Segmented genome. An unusual feature of influenza virus is its segmented genome (**Fig. 11.13B**). The influenza A genome includes eight segments, each a separate linear (–) strand of RNA. The segmented genome consists of multiple separate nucleic acids, like the multiple chromosomes of a eukaryotic cell. Within the virion, each RNA segment forms a loop, complexed with NP subunits that condense the RNA in a helix. (The helix does not involve base pairs, and it differs from the standard helical forms of RNA.) The two ends of RNA are complexed with the RNA-dependent RNA polymerase, poised for RNA synthesis early in infection.

Given that viral infection requires all eight segments, how does the assembly mechanism package exactly eight segments, one of each? Cryo-electron tomography shows how the eight segments package precisely within the virion: The segments appear to link to each other in order as they arrange themselves (**Fig. 11.13B)** . The computed images show sections across the influenza virion, in which all eight RNA segments stand side by side, like a bundle of sticks. Sections taken from different depths through the particle reveal tiny molecular connections between adjacent segments. Further experiments with genetic constructs and fluorescence microscopy confirm that all eight unique segments link together in a defined, reproducible pattern.

A segmented genome has profound consequences for viral evolution. If two different strains of influenza virus infect a host simultaneously, their segments can reassort to generate a novel hybrid strain. Because influenza genomes are capable of reassortment , they can rapidly generate a new strain that our immune system fails to recognize, such as the pandemic H1N1 strain of 2009 (discussed shortly).

Note: Distinguish between reassortment (two different viruses contribute separate genome segments to a reassortant genome) and recombination (two different viruses contribute genetic material to a recombinant molecule.)

Within a virion, the NP-coated RNA segments (or RNPs) are loosely contained by a shell of matrix proteins (M1) (**Fig. 11.14**). The matrix layer is further enclosed by the envelope. The envelope derives from the phospholipid membrane of the host cell, which incorporates the viral glycoproteins hemagglutinin (HA) and neuraminidase (NA). Upon host cell infection, the viral HA and NA bind specific carbohydrate chains on host cell-surface glycoproteins; the chains end in sialic acid. When a newly formed virion exits its host cell, neuraminidase acts as an enzyme to cleave a cell membrane glycoprotein, thus releasing the virion outside the cell. Neuraminidase can be blocked by the antiviral agent oseltamivir (Tamiflu), one of the main drugs available to treat influenza.

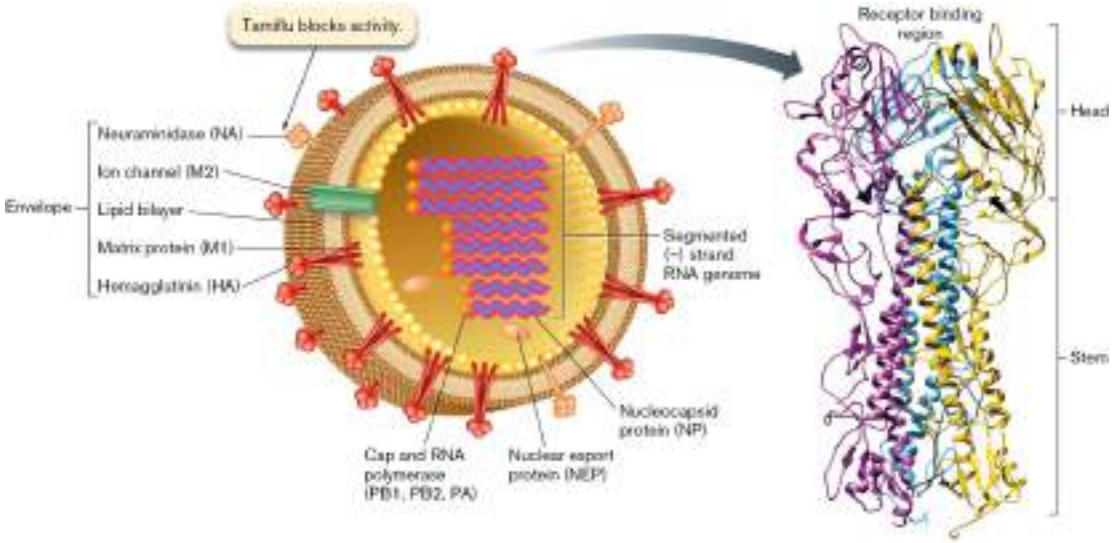


FIGURE 11.14 ■ Structure of influenza. Diagram of influenza A virion structure, showing envelope (colored tan), envelope proteins, matrix protein (yellow), RNA segments (blue) with nucleocapsid proteins (red) and attached polymerase, and

the nuclear export protein (NEP). The brush-like border coating the envelope consists of glycoproteins, hemagglutinin (HA), and neuraminidase (NA). **Inset:** Hemagglutinin trimer complex, molecular structure.

How does the virus express genes encoded by its (−) strand RNA genome? A complementary (+) strand RNA must be synthesized within the infected cell. Within the infected cell, each (−) strand segment must be transcribed to a (+) strand mRNA having a host-derived 5′ cap and viral RNA polymerase and a 3′ poly-A tail (AAA-OH ^{3} ′) typical of eukaryotic mRNA. Thus, within the virion each NP-coated RNA segment carries its own RNA-dependent RNA polymerase complex (proteins PB1, PB2, PA; **Figs. 11.13** and **11.14**). The polymerase proteins were expressed in the previous host, before the virions formed and the cell lysed.

Once the mRNA molecules are synthesized, they are ready for processing (including 5′ cap and 3′ tail) and translation by host ribosomes. Certain segments (1, 2, 7, 8) encode more than one type of product from a common sequence (**Fig. 11.15**). The multiple products are enabled by alternate mRNA processing and by ribosome slippage and frameshifting. These nonstandard expression mechanisms are typical of the way RNA viruses economize to make diverse products from a small genome.

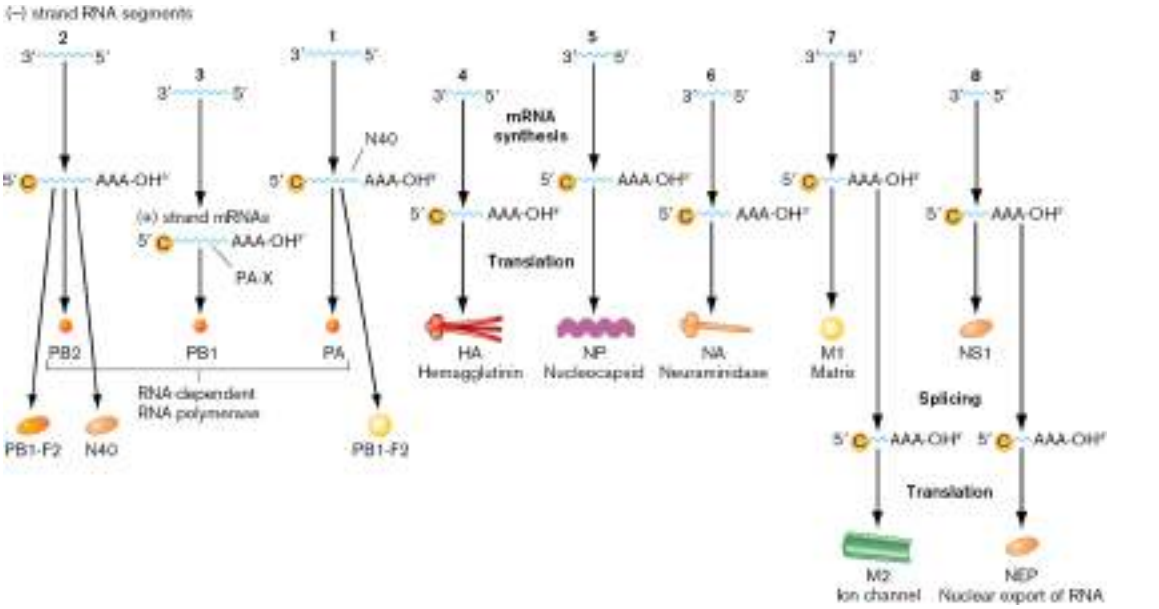


FIGURE 11.15 ■ Genome of influenza A. The influenza A genome consists of eight RNA segments, each encoding one or two proteins. Each (−) strand segment must be transcribed to a (+) strand mRNA having a host-derived 5′ cap and viral RNA polymerase and a 3′ poly-A tail (AAA-OH ^{3} ′) consisting of a variable number of adenine nucleotides with a 3′ hydroxyl end. Segments 1, 2, 7, and 8 are processed to express multiple proteins.

Certain influenza A proteins interact with the host cell proteome and alter the host defenses. For example, protein NS1 suppresses components of the host innate immunity (host defenses that are built in, not adaptive; discussed in Chapters 23 and 24). The innate immunity components suppressed by NS1 include type 1 interferon and RNA-activated protein kinases. This inhibition of host defenses allows IAV particles to infect lung epithelial tissues.

Figure 11.16 shows results of an experiment by Eike Hrincius and colleagues at the University of Muenster, Germany. Mice were infected with IAV, and their lung tissues were sectioned. Hematoxylin stained the cell nuclei dark blue, revealing the dense packing of the epithelial cells lining the bronchioles (passages that carry air to the alveoli). Virus infection (brown) was revealed by

immunohistochemistry , a technique that first reacts an animal antibody with the virus, followed by a secondary antibody that binds the first and generates a stain. In this case, a goat anti-IAV antibody was used with a biotinylated secondary antibody that generates a streptavidin–horseradish peroxidase reaction, which stains brown. The IAV-infected mouse lung shows brown stain lining the bronchioles (**Fig. 11.16** , middle panel). But the control and the IAV with an NS1 mutation show much less infection, because the host defenses are not inhibited by NS1.

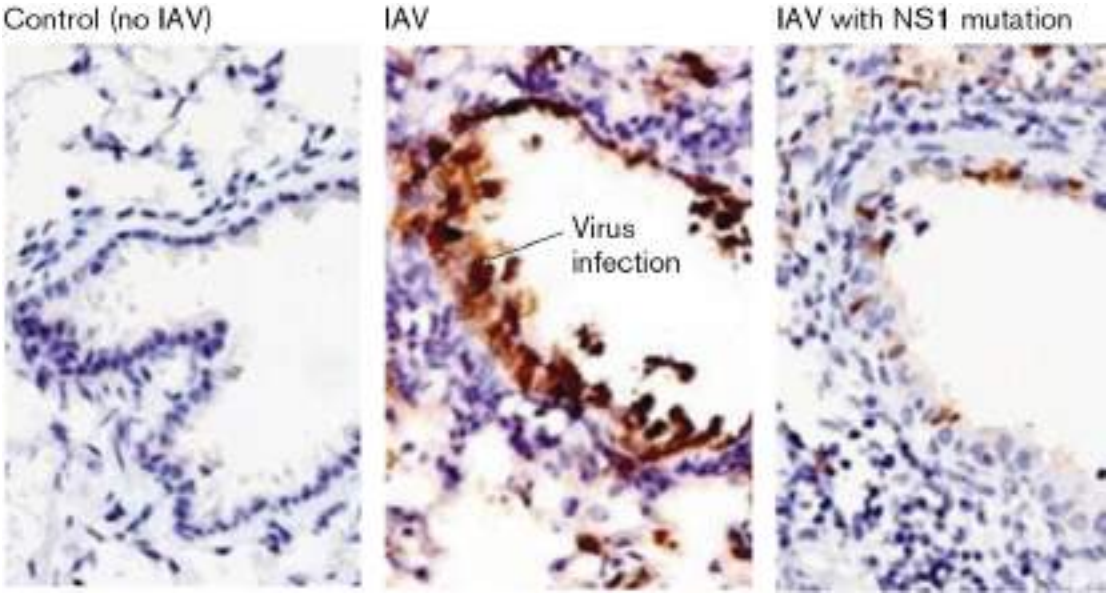


FIGURE 11.16 ■ IAV infection of bronchiole epithelium requires NS1 protein. Mice were infected with virus and their lungs sectioned with hematoxylin stain (cell nuclei are stained blue). Immunostain with anti-IAV antibodies reveals virus infection (brown). Bronchioles are shown at 40× magnification (light microscopy).

E. R. HRINCIUS ET AL. 2012. *AM J PATHOL* . **180** :2361–74

The immunohistochemistry shows the specific tropism of influenza virus for epithelial cells. Tropism is important because the specific tissues infected can determine the symptoms and consequences of infection. For example, a different respiratory virus,

the coronavirus SARS-CoV-2, binds the ACE2 receptor found in vascular tissues. The coronavirus infects vascular tissues of not only the lung but also the intestine, heart, and brain. Thus, SARS-CoV-2 has a broader range of symptoms and sequelae (long-term effects) than does influenza virus.

Mutation and reassortment. Influenza viruses gradually accumulate mutations leading to newly virulent strains that go unrecognized by the host immune system. This gradual accumulation of mutations is known as "antigenic drift." Antigenic drift allows influenza viruses to continually acquire small mutations that can lead to new phenotypes with respect to drug resistance and host range.

A more radical way that strains change is reassortment. The key advantage of a segmented genome is that it enables reassortment of segments between two strains coinfecting the same cell. Reassortment leads to "antigenic shift," in which a new strain combines a mixture of components derived from two different strains. The new strain may enter a new host, where it evades the host immune system. Even strains that infect animals such as ducks or swine may reassort their segments with those of a coinfecting human virus.

Influx of genes from a distantly related strain can sharply increase virulence and mortality. For example, the 1968 Hong Kong flu strain, which killed over 33,000 people in the United States, derived three segments from avian strains. Major epidemics of exceptionally virulent influenza arise as a result of reassortment with genome segments from strains that evolved within ducks or swine, agricultural animals that live in close proximity to humans. In each genome, the "H" and "N" numbers designate alleles of the genes encoding envelope proteins h emagglutinin and n euraminidase, respectively. For example, the Hong Kong flu strain had alleles H3 and N2 (which have since become prevalent in "seasonal" human-flu strains).

In 1979, in Europe, an avian flu strain was found to have "jumped" into swine (the "avian-like" swine strain). In 1992, a

triple-reassortant strain was identified that included segments PB2 and PA (encoding RNA-dependent RNA polymerase) from an avian virus; PB1 (polymerase subunit), NP (nucleocapsid), and M (matrix protein) from a swine virus; and PB, H, and N from a human seasonal strain of influenza A. Because the H and N envelope proteins came from a human strain, the triple reassortant could be transmitted readily between humans. Today, molecular surveillance reveals many emerging strains of human influenza A that combine avian and swine alleles. The "swine" strain in 2009 had alleles H1 and N1, similar to the 1918 pandemic strain (**Fig. 11.17**). In 2009, prompt public health measures such as quarantine helped keep the disease rate low. Yet another reassortant avian strain, H7N9, emerged in China in 2013. So far, these avian strains have been contained—but the next time we may be less fortunate.

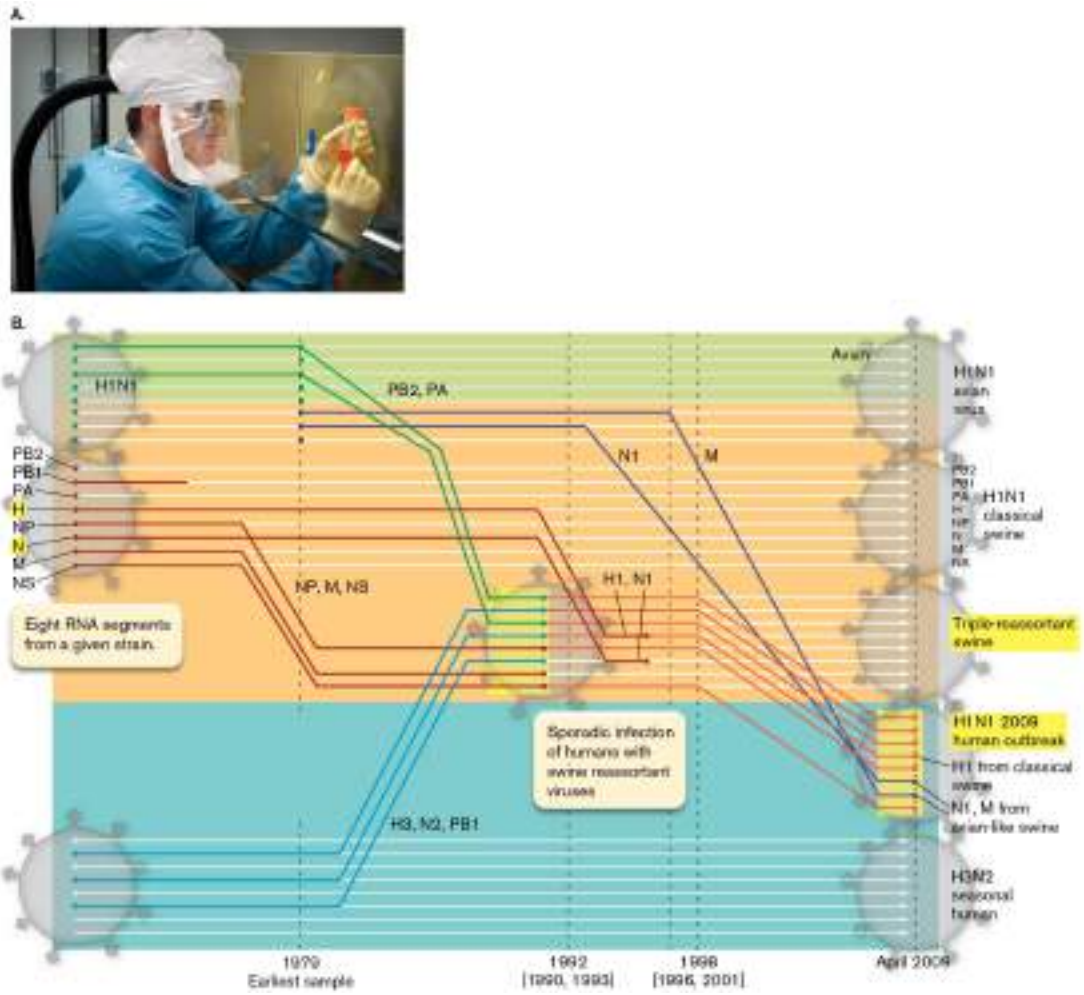


FIGURE 11.17 ■ Reassortment between human, avian, and swine strains generates exceptionally virulent strains of influenza A. A. The reconstructed strain of the 1918 pandemic influenza virus is studied by Terrence Tumpey, a microbiologist at the Centers for Disease Control and Prevention (CDC), Atlanta. **B.** The 2009 strain of H1N1 influenza A arose from a series of reassortments of the eight RNA segments from avian, swine, and human influenza strains. The numbers following "H" and "N" refer to different alleles of the genes encoding hemagglutinin and neuraminidase, respectively.

Source: Part B modified from Gavin J. D. Smith et al. 2009. *Nature* **459** :1122.

CDC/JAMES GATHANY

How will the cycle of influenza pandemics respond to climate change? Climate modelers attempt to address this question by studying past pandemics in the context of weather patterns. Many factors are involved, from the behavior of animal populations to the shifting trends in temperature and humidity. While the models are complex, there is growing evidence that influenza pandemics will increase with increased rapid changes in weather.



Thought Question

11.5 Could the influenza genome change by recombination of segments, rather than by reassortment? What about the lambda phage genome?

Transmission and Attachment to Host Cell

What determines which animals can be infected by a given flu strain? One factor is the requirement for a host cell protease to cleave the hemagglutinin protein on the virion envelope. Cleavage of hemagglutinin enables a small peptide called the fusion peptide to mediate viral entry into the host cell (discussed shortly). The presence of the protease is one host factor that determines which kind of host may be infected and which tissues within the host support viral replication. Host factors of many kinds mediate viral infections.

Influenza receptor is a sialic acid glycoprotein. Another important host factor for influenza is cell-surface glycoproteins that contain a terminal sialic acid (**Fig. 11.18**). The sialic acid polysaccharide of the glycoprotein binds hemagglutinin, attaching the virion and enabling endocytosis. The precise structure of the sialic acid host receptor may determine whether a strain such as avian influenza H5N1 will spread directly between humans. For example, the sialic acid connection in the receptor polysaccharide can involve different OH groups of the sugar galactose: a linkage to the OH-3 (alpha-2,3) or to the OH-6 (alpha-2,6) bond (**Fig. 11.18**). The influenza strain H5N1 recognizes mainly the alpha-2,3-linked protein, found in birds. In humans, the upper respiratory tract contains mainly alpha-2,6-linked receptors; alpha-2,3-linked receptors are found only deeper within the lungs. But swine carry receptors of both types. For this reason, swine are believed to act as a "mixing bowl" for strains from birds and humans, as well as from swine. Thus, swine incubated the avian strain in 1979 and then enabled later reassortment with human-flu genome segments, leading eventually to the 2009 strain—which sickened both pigs and humans. Today, large swine facilities are monitored for appearance of novel reassortant strains.

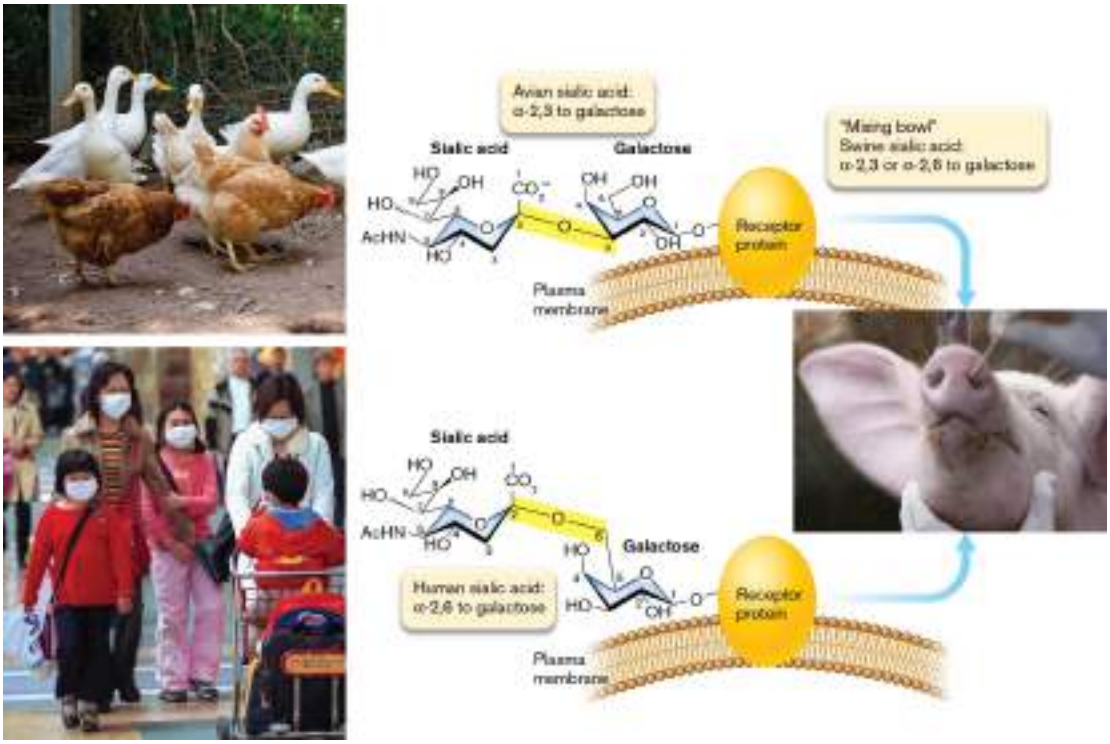


FIGURE 11.18 ■ Influenza receptors in different hosts. The avian host receptor polysaccharide contains sialic acid with an alpha-2,3 bond to galactose, whereas the human receptor in the upper respiratory tract has an alpha-2,6 bond. Swine receptors include both forms of sialic acid; thus, swine can be infected by both avian and human strains and may act as a "mixing bowl" for reassortment.

CHRISTOPHER MORRIS/CORBIS VIA GETTY IMAGES
ZERO CREATIVES/CULTURA/CORBIS
GETTY IMAGES/DIMAS ARDIAN

The avian influenza strain H5N1 causes exceptionally high mortality in humans, but it is rarely transmitted from one person to another. More rapid transmission could arise from antigenic drift, by accumulating mutations in the gene encoding avian hemagglutinin. Rapid person-to-person transmission of H5N1 might cause an influenza pandemic with high mortality. That is why public health organizations were so concerned when, in 2011, researchers announced that they had identified mutations conferring H5N1

transmission in the ferret model system, which closely resembles the human system (Special Topic 11).

Thought Question

11.6 How could swine play a role in generating a pandemic strain of influenza? How could avian influenza strain H7N9 become a pandemic strain endangering many people?

Host Entry and Replication Cycle

For replication, the influenza virus must gain entry to the host cell and evade its molecular defenses. Then the viral components travel in and out of the nucleus. Several viral enzymes and structural proteins travel along with the virus's genetic material. As progeny envelope proteins are made, they require transport through the endoplasmic reticulum (ER) and the Golgi complex to the cell membrane. The overall replication cycle is highly complex—and the molecular details offer many opportunities to devise antiviral agents.

Endocytosis and membrane fusion. Endocytosis of the influenza virion involves a key step of acid-mediated membrane fusion, which offers a target for antiviral agents (**Fig. 11.19**). As the influenza virion binds its sialic acid receptor (**Fig. 11.19** , step 1), a host protease cleaves each HA, forming a fusion peptide (step 2). The hemagglutinin trimer now contains three N-terminal fusion peptides. A fusion peptide is a portion of an envelope protein (cleaved from hemagglutinin in the case of influenza virus) that changes conformation so as to facilitate envelope fusion with the host cell membrane.

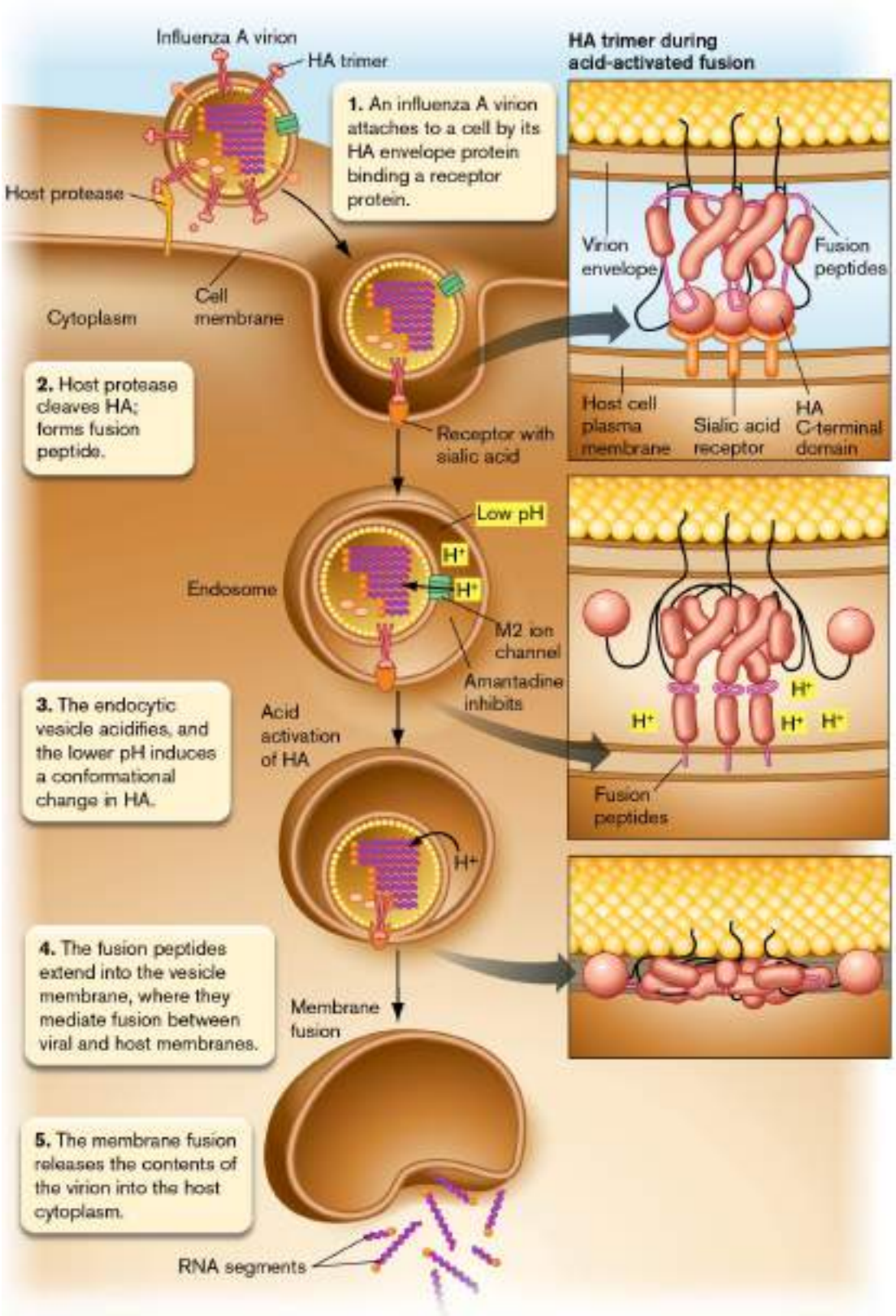

Influenza A virion
HA trimer
Host protease
1. An influenza A virion attaches to a cell by its HA envelope protein binding a receptor protein.
Cytoplasm
Cell membrane
2. Host protease cleaves HA; forms fusion peptide.
Receptor with sialic acid
Low pH
H+
H+
Endosome
M2 ion channel
Amantadine inhibits
3. The endocytic vesicle acidifies, and the lower pH induces a conformational change in HA.
Acid activation of HA
H+
4. The fusion peptides extend into the vesicle membrane, where they mediate fusion between viral and host membranes.
Membrane fusion
5. The membrane fusion releases the contents of the virion into the host cytoplasm.
RNA segments
HA trimer during acid-activated fusion
Virion envelope
Fusion peptides
Host cell plasma membrane
Sialic acid receptor
HA C-terminal domain
H+
H+
H+
H+
Fusion peptides

FIGURE 11.19 ■ Influenza virion attachment to receptor, acid activation, and release of genome in the cytoplasm.

When the virion is taken up by endocytosis, the endocytic vesicle fuses with a lysosome and its interior acidifies (**Fig. 11.19** , step 3). The lowered pH (increased H^+ concentration) drives H^+ ions into the virion through the M2 ion channel in the matrix layer (see **Figure 11.14** for virion structure). The influx of acid causes the matrix proteins to dissociate from the NP-coated RNA. The M2 ion channel is the target of amantadine, one of the first anti-influenza drugs; unfortunately, most strains today have evolved resistance to the drug. Low pH also induces a conformational change, shifting the C-terminal ends back and the N-terminal fusion peptides outward to face the vesicle membrane. The peptides extend into the membrane (step 4), where they mediate fusion between viral and host membranes. The fusion process expels the contents of the virion into the host cytoplasm (step 5).

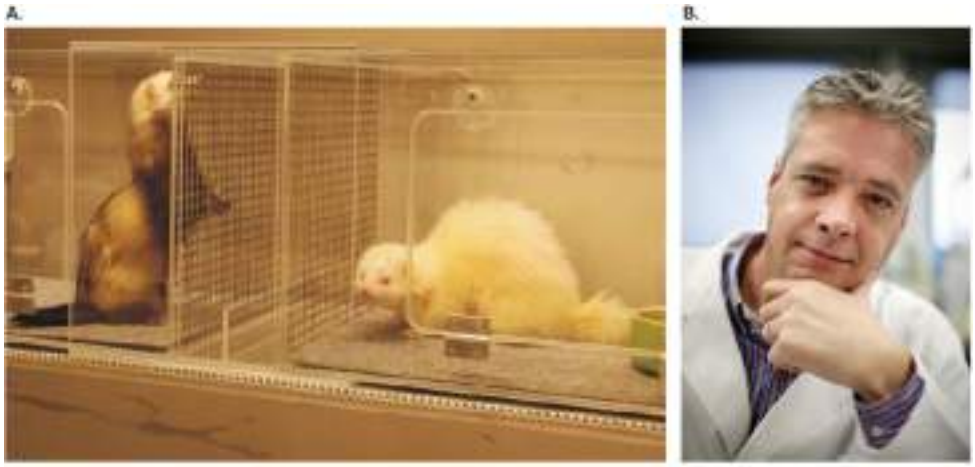
SPECIAL TOPIC 11 Designing a Pandemic Flu

Pandemics usually arise from an emerging pathogen—a virus or bacterium that rarely infects humans but suddenly spreads widely. Such change requires mutations that increase transmission. With the coronavirus SARS-CoV-2, scientists wondered how a deadly virus with poor transmission could have mutated to transmit so readily. If we had known, could we have prevented the COVID-19 pandemic? And could we prevent future coronavirus or influenza pandemics? The answer might be yes. But how could we safely perform the

experiments to find out? Might our experiments actually start a pandemic?

Research to test how an emerging strain becomes a pandemic threat is called "gain-of-function." Gain-of-function research involves the serial passaging of a pathogen from host to host, under selection pressure for mutations that increase transmissibility, virulence, and host tissue tropism. Such research on coronaviruses and influenza viruses is considered essential, but highly dangerous—and, thus, debated in the research community and in government regulatory agencies.

A gain-of-function experiment was conducted on influenza virus by Ron Fouchier, at Erasmus University Rotterdam, with funding from the European Union and from the United States National Institutes of Health. In 2012, Fouchier and colleagues conducted experiments to explore the increase of transmissibility of influenza A in ferrets (**Fig. ST 11.1A**). Ferrets have an upper respiratory tract that shows many features in common with the human tract (see also Special Topic 6). Thus, ferrets are considered a good model system for human influenza transmission.



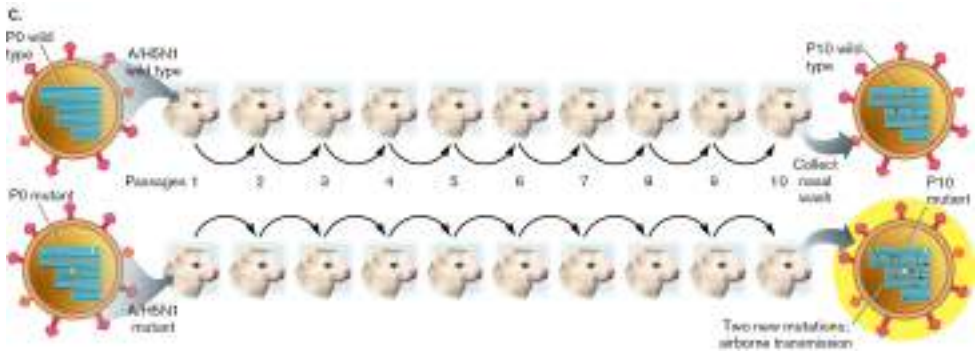


FIGURE ST 11.1 ■ Influenza transmission in ferrets. **A.** Ferrets in cages that allow airflow. **B.** Ron Fouchier. **C.** Intranasal passage of a mutant influenza virus selects for mutations that enable airborne transmission. *Source:* Part C modified from S. Herfst et al. 2012. *Science* **336** :1534–1521, fig. 2.

S. HERFST ET AL. 2012. *SCIENCE* **336** :1534. REPRINTED WITH PERMISSION FROM AAAS

DIRK-JAN VISSER/THE NEW YORK TIMES/REDUX

Fouchier (**Fig. ST 11.1B**) tested two variants of H5N1, the "avian" strain that shows high lethality but low transmission in humans. One H5N1 variant was engineered to contain four point mutations known to increase transmission of other strains. For example, two of the mutations increased the binding of hemagglutinin to human receptors. Initially, both variants showed no airborne transmission in ferrets. The ferrets were kept in separate cages, allowing no direct contact. But Fouchier passaged each H5N1 variant intranasally; that is, from nose to nose of infected ferrets (**Fig. ST 11.1C**). After ten intranasal passages (P1–P10 in **Figure ST 11.1C**), the engineered variant suddenly showed airborne transmission. The new phenotype was associated with two extra mutations. The implication of this result was that a small number of specific mutations might convert avian influenza into a pandemic strain.

Fouchier's work generated enormous controversy, including a lawsuit to prevent publication of key details and a halt to U.S. government funding of certain projects. Despite the biosafety precautions (**Fig. ST 11.2**), what if such research could actually lead to escape of a pandemic strain? Does the knowledge gained actually aid prevention or is it useless, given that other unknown mutations could still enhance transmission? Ultimately, a moratorium was declared on gain-of-function research until stricter biosafety regulations were instituted. But the debate continues.



FIGURE ST 11.2 ■ Biosafety level 3 laboratory. Study of influenza transmission requires special precautions.
VÉRONIQUE BURGER/SCIENCE SOURCE

RESEARCH QUESTION

What questions could be tested about the evolution of pandemic influenza or coronavirus? Are these experiments worth the risk?

Herfst, Sander, Eefje J. A. Schrauwen, Martin Linster, Salin Chutinimitkul, Emmie de Wit, et al. 2012. Airborne transmission of influenza A/H5N1 virus between ferrets. *Science* **336** :1534–1541.

Belser, Jessica A., Wendy Barclay, Ian Barr, Ron A. M. Fouchier, Ryota Matsuyama, et al. 2018. Ferrets as models for influenza virus transmission studies and pandemic risk assessments. *Emerging Infectious Diseases* **24** :965–971.

Synthesis of (+) strand mRNA. After the influenza virion is taken up by endocytosis (**Fig. 11.20** , step 1), all the viral (−) RNA segments are released in the cytoplasm (step 2). Each RNA retains its coat of nucleocapsid proteins, as well as a prepackaged RNA-dependent RNA polymerase. The NP-coated RNA segments individually pass through a nuclear pore into the nucleus (step 3). Within the nucleus, each genomic (−) RNA segment with its prepackaged polymerase synthesizes (+) strand RNA for mRNA (step 4). Each mRNA synthesis initiates with a 7-methylguanosine "capped" RNA fragment (portrayed as a 5′ "C" in **Fig. 11.20**). The influenza polymerase obtains the cap fragments from the host by cleaving them from host nuclear pre-mRNA—a process quaintly known as "cap snatching." The (+) strand mRNA molecules return to the cytoplasm for translation (step 5), using the snatched cap to bind the host ribosome. The RNA segments encoding envelope proteins attach to the ER for protein synthesis and transport to the host cell membrane (step 6). The newly synthesized nucleocapsid proteins (NPs), as well as RNA-dependent RNA polymerase components, subsequently return to the nucleus (step 7). Other genome-packaging proteins (M1 and NEP/NS2) also return to the nucleus.

Synthesis of (+) strand and (−) strand genomic RNA. Back in the nucleus, the original (−) strand RNA segments also serve as templates for RNA synthesis, without cap snatching (**Fig. 11.20** ,

step 8). The uncapped (+) strand RNA then becomes coated with the newly made NP subunits imported from the cytoplasm. The NP-coated (+) strand serves as a template to synthesize (−) strand RNA genomes (step 9), which also become coated with NP (step 10).

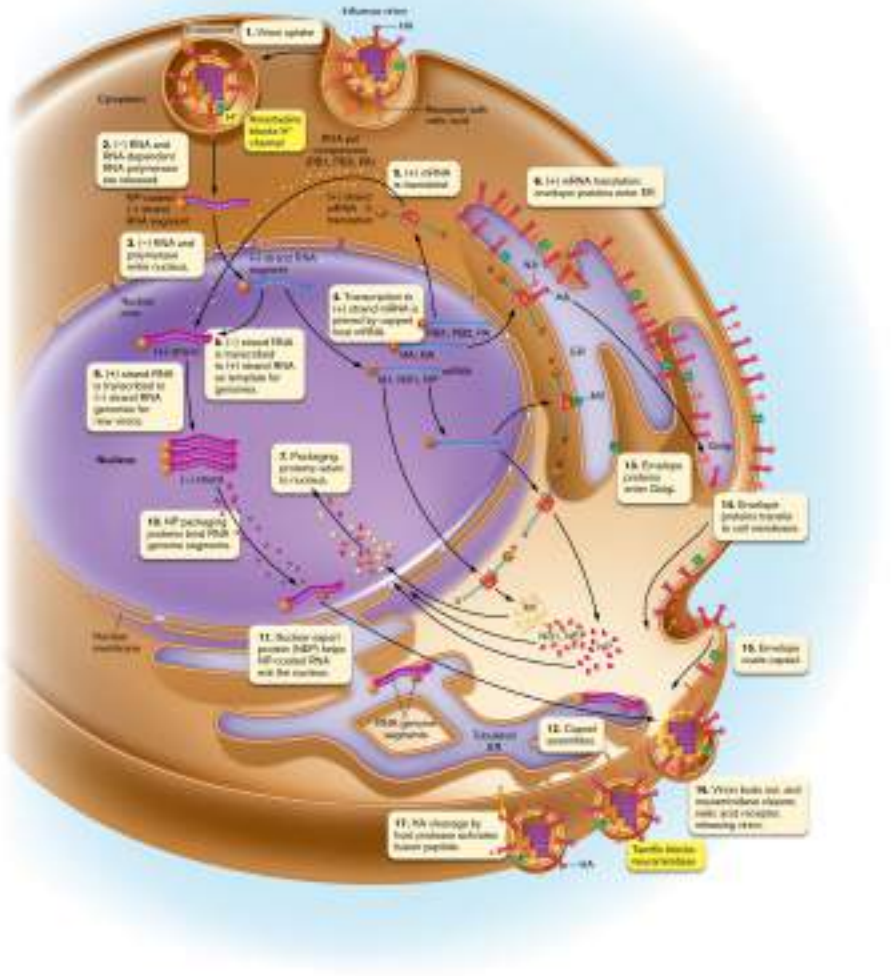


FIGURE 11.20 ■ Replication of influenza virus.

The NP-coated (−) RNA associates with a newly made polymerase for a future cycle of viral replication. The RNA is then complexed with matrix protein (M1) and nuclear export protein (NEP)—proteins that were imported from the cytoplasm earlier. At

last, the fully packaged (−) RNA segments exit the nucleus to the cytoplasm (**Fig. 11.20** , step 11), where they approach the cell membrane for packaging into progeny virions (step 12).

Envelope synthesis and assembly. The envelope proteins synthesized at the ER include hemagglutinin (HA) and neuraminidase (NA). Within the ER lumen, these proteins are glycosylated by host enzymes and then transferred to the Golgi (**Fig. 11.20** , step 13) for export to the cell membrane (step 14). Within the cell membrane, the envelope proteins assemble around a group of (−) RNA segments complexed with their matrix and packaging proteins, completing the virion particle (step 15).

To exit the cell, the virion buds out (**Fig. 11.20** , step 16). Viral exit requires a final step of host release: Neuraminidase (the envelope protein NA) cuts the sialic acid link of the host glycoproteins (step 17), releasing the virion out into the bloodstream. This host release activity of neuraminidase is inhibited by oseltamivir (Tamiflu), the main antiviral agent currently useful against influenza. Tamiflu remains useful today, but resistant strains of the 2009 H1N1 virus have emerged, so we urgently need new antivirals ahead of the next influenza pandemic.

Experimental evidence. How do researchers figure out all the steps of replication shown in **Figure 11.20** and then identify targets for drugs? A key technique is fluorescence microscopy (described in Chapters 2 and 3), along with the related phenomenon of bioluminescence, in which a biochemical reaction leads to light emission. For example, luciferase enzyme hydrolyzes ATP to emit green light. Fluorescence and bioluminescence are highly sensitive tools that reveal specific parts of cells and viral factories.

Figure 11.21 shows an example of an experiment by Richard Kao and colleagues at the University of Hong Kong. Their experiment tested the ability of the drug nucleozin to block entry of viral nuclear packaging protein (NP) into the nucleus (step 7, **Fig. 11.20**). In their experiment, cultured animal cells were infected

with IAV in the presence or absence of nucleozin. DAPI fluorescence labels the nucleus, and luciferase-conjugated antibodies label the NPs. In the presence of nucleozin, the DAPI-stained nucleus remains blue, excluding the green color of NP. Without nucleozin, NP enters the nucleus, which then combines the blue and green colors, appearing cyan. Such experiments reveal not only the steps of replication but also the targets for promising antiviral agents of the future.

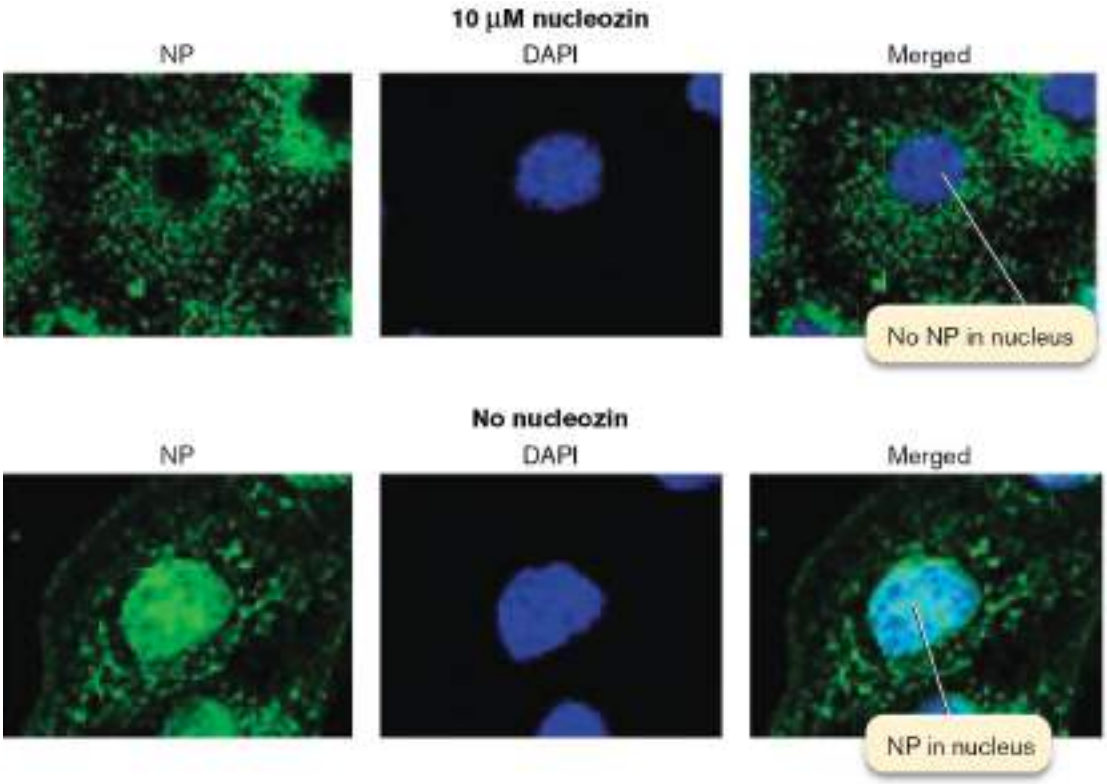


FIGURE 11.21 ■ The drug nucleozin inhibits NP entry into the nucleus. Nuclei are stained with DNA-binding fluorophore 4′, 6-diamidino-2-phenylindole (DAPI; blue). NP was localized with antibody label and luciferase (green). Colocalization of NPs in the nucleus appears cyan.

R. KAO ET AL. 2010. *NAT BIOTECHNOL* . **28** :600–605
R. KAO ET AL. 2010. *NAT BIOTECHNOL* . **28** :600–605

To Summarize

- **Influenza virus** causes periodic pandemics of respiratory disease. New virulent strains arise via antigenic drift (the gradual accumulation of mutations) and more suddenly by antigenic shift (through reassortment of gene segments from very different strains, such as human, avian, and/or swine strains).
- **The influenza virus consists of segmented (−) strand RNA.** Each segment is packaged with nucleocapsid proteins. Segments from different strains reassort through coinfection of a shared host cell.
- **Nucleocapsid and matrix proteins** enclose the RNA segments of the influenza virus. The matrix is surrounded by an envelope containing the glycoproteins hemagglutinin (HA) and neuraminidase (NA).
- **Immunohistochemistry** reveals the presence of invading virus in tissue and the tropism of the virus (tissue preference for infection).
- **Envelope HA proteins mediate virion attachment.** The HA protein includes a fusion peptide that undergoes conformational change to cause fusion between the viral envelope and host cell membrane. For influenza, the virion is internalized by endocytosis.
- **Lysosome fusion with endosomes** triggers viral envelope fusion with the endosome membrane. The viral genome and proteins are then released into the cytoplasm. Viral (−) strand RNA segments attached to RNA-dependent RNA polymerase enter the nucleus.
- **Influenza mRNA synthesis initiates with a capped RNA fragment** cleaved from host mRNA. The capped viral mRNAs return to the cytoplasm for translation.
- **Radioisotope fluorophores can label specific viral and host cell proteins involved in viral infection.**

- **Genomic RNA synthesis generates a (+) strand RNA as a template for (−) strand RNA segments.** Progeny RNA segments are then packaged in newly made nucleocapsid protein and exported to the cytoplasm for coating with matrix, host cell membrane, and viral envelope proteins.
- **Neuraminidase cleaves the sialic acid connection to host glycoproteins.** This key step of virion release can be blocked by oseltamivir (Tamiflu).

Glossary

nucleocapsid protein (NP)
: A protein that coats a viral genome.

segmented genome
: A viral genome that consists of more than one nucleic acid molecule.

reassortment
: The packaging of viral chromosome segments from two different viruses into one progeny virion. Refers to separate segments from a segmented genome, without helix recombination.

reassortment
: The packaging of viral chromosome segments from two different viruses into one progeny virion. Refers to separate segments from a segmented genome, without helix recombination.

recombination
: See also *homologous recombination* . The process by which two DNA molecules exchange arms by cutting and splicing their helix backbones.

matrix protein
: A protein, found in some viruses, that is located between the capsid and the membrane envelope.

immunohistochemistry
: The use of a labeled antibody to stain and visualize specific structures in tissue.

fusion peptide
: A portion of a viral envelope protein that changes shape to facilitate envelope fusion with the host cell membrane.

host factor
: A trait of an individual host that affects susceptibility to disease, in comparison with other individuals.

endocytosis
: The invagination of the cell membrane to form a vesicle that contains extracellular material.

11.3 Human Immunodeficiency Virus (HIV): Retrovirus

Human immunodeficiency virus (HIV) is a retrovirus (Baltimore group VI). As discussed in Chapter 6 , a retrovirus requires the enzyme reverse transcriptase (RT) to copy its RNA genome into DNA. The most famous lentivirus is human immunodeficiency virus (HIV) , the cause of acquired immunodeficiency syndrome (AIDS) . Today, HIV remains a global threat. In 2020, according to the United Nations, 38 million people globally were living with HIV, and AIDS-related infections claimed two-thirds of a million lives. The rate of new infections has declined, thanks to research and new drugs that can save lives and prevent HIV transmission. But these therapies need to reach more of the world's population.

Retroviruses are a large family of viruses known to infect all types of vertebrate and invertebrate animals; for examples, see **Table 11.1** . The "simple retroviruses" have genomes of just four genes, such as feline leukemia virus (FeLV), the number one killer of outdoor cats in the United States. Simple retroviruses generally cause cancer. The lentiviruses , or "slow viruses," cause diseases that progress slowly over many years. Lentiviruses possess additional regulator genes that modulate host interactions.

TABLE 11.1	**Retroviruses of Animals (Examples)**		
Genus	**Virus**	**Disease(s)**	**Hosts**
Simple retroviruses			

TABLE 11.1	Retroviruses of Animals (Examples)		
Alpharetrovirus	Avian leukosis virus (ALV)	Leukemia	Birds
	Rous sarcoma virus (RSV)	Sarcoma (tumor)	Birds
Betaretrovirus	Mouse mammary tumor virus (MMTV)	Mammary tumor	Mice
Gammaretrovirus	Feline leukemia virus (FeLV)	Lymphoma, immunodeficiency	Cats
	Moloney murine leukemia virus (MMLV)	Leukemia	Mice
Deltaretrovirus	Bovine leukemia virus (BLV)	Leukemia	Cattle
	Primate T-lymphotrophic virus (PTLV-1) [formerly human T-cell leukemia virus (HTLV)]	Leukemia	Humans
Epsilonretrovirus	Walleye dermal sarcoma virus	Sarcoma	Fish

TABLE 11.1	Retroviruses of Animals (Examples)		
	(WDSV)		
Lentiviruses			
Lentivirus	Human immunodeficiency virus (HIV-1, HIV-2)	AIDS	Humans
	Simian immunodeficiency virus (SIV)	Simian AIDS	Monkeys
	Equine infectious anemia virus (EIAV)	Anemia	Horses
	Maedi-Visna virus (MV)	Neurological disease	Sheep

In addition, most animal genomes show evidence of endogenous retroviruses , sequences from an ancient retrovirus whose genome integrated and became "fixed" by mutation, such as HERV-K (discussed in Section 11.4). Surprisingly, endogenous retroviruses can evolve into essential parts of host genomes. And the human immunodeficiency virus has now been engineered to make "lentivectors," our most successful agents of gene therapy (see Sections 11.4 and 16.6).

History of HIV and AIDS

HIV is a lentivirus that evolved from viruses infecting African monkeys. Two major types are recognized: HIV-1, the cause of most infections at present; and HIV-2, which appears to have evolved independently from a different strain infecting monkeys. The virus is transmitted through blood and through genital or oral-genital contact. HIV can hide in the host cell for many years, with only gradual buildup of virus particles, most of which are eliminated by the host. Eventually, however, the virus destroys the body's T lymphocytes, leaving the host defenseless against many organisms that normally would be harmless. The progression from HIV infection to AIDS disease is described in Section 26.1 .

When AIDS first emerged in the 1980s, governments were slow to address the disease. American society failed for many years to grasp the significance of AIDS because the syndrome first appeared in societal groups considered marginal (homosexual men and certain ethnic immigrants), although it spread to all social classes. A major role in AIDS awareness was played by the arts, such as the AIDS Memorial Quilt, an ongoing project documenting the lives of over 100,000 persons, and the 1993 film *And the Band Played On* , based on the book by Randy Shilts documenting society's failure to address the disease. AIDS patients and activist supporters staged many public demonstrations, including an unprecedented protest at the National Institutes of Health to demand increased funding for research. Afterward, the NIH Office of AIDS Research accelerated progress on AIDS research, funding, for example, the discovery of antiviral protease inhibitors (discussed later).

At first the AIDS virus proved extremely difficult to detect and grow in culture. The virus HIV and its causative role in AIDS were discovered in 1983 by French virologist Luc Montagnier (1932–2022), building on Robert Gallo's studies of retroviruses (**Fig. 11.22A)** . In 2008, the Nobel Prize in Physiology or Medicine was awarded to Montagnier and Françoise Barré-Sinoussi (**Fig. 11.22B**) for their discovery of HIV and its role in AIDS. Since that time, scientists from many countries have collaborated to develop therapies that prevent the symptoms of AIDS and transmission of HIV-1. The different antiretroviral drugs target different molecular mechanisms of HIV infection, as described in this chapter. For those with access to

medicine, antiretroviral therapy (ART), a mixture of antiretroviral drugs, can enable people carrying HIV to lead a normal life. But infected individuals must take these expensive drugs indefinitely, and physicians must check their patients to make sure the virus does not become resistant.

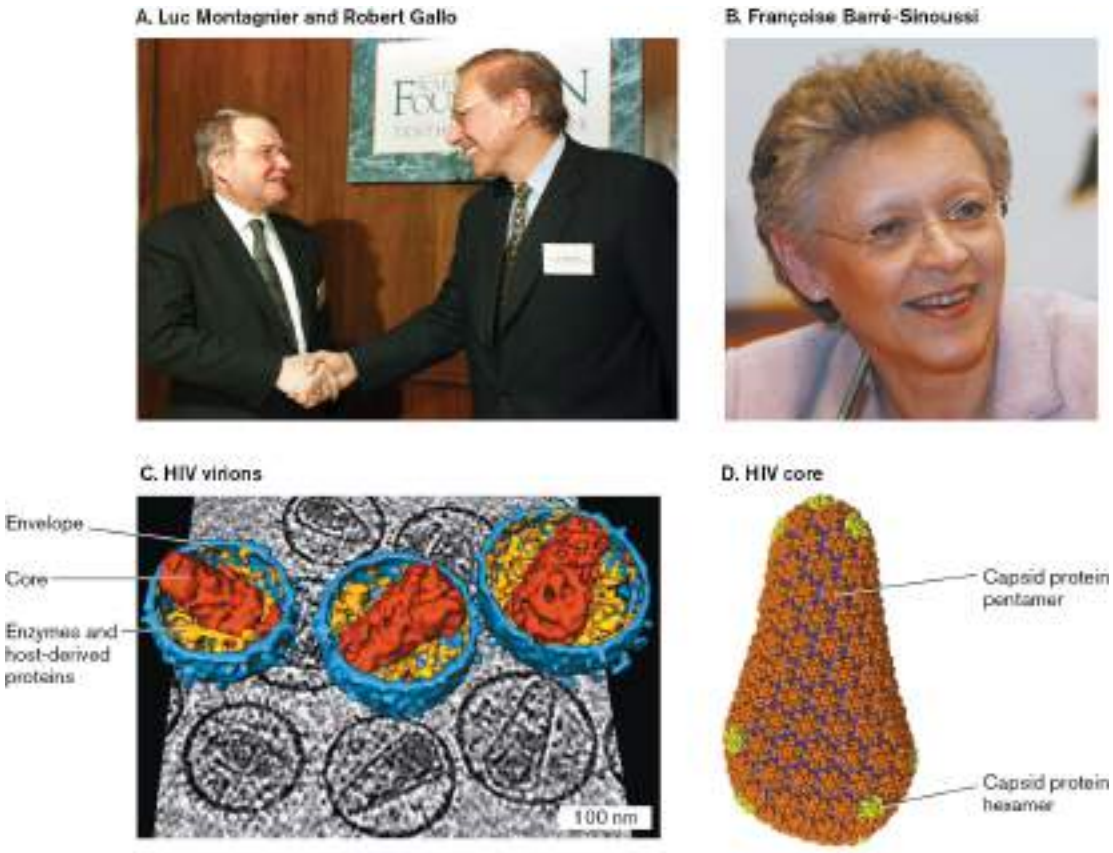


FIGURE 11.22 ■ HIV discovery. A. Luc Montagnier (left) and Robert Gallo agree to collaborate on development of an AIDS vaccine, 2002. **B.** Françoise Barré-Sinoussi, at the Pasteur Institute, worked with Montagnier to discover the virus that causes AIDS. **C.** HIV virions (cryo-EM tomography). **D.** HIV core, model from cryo-EM and X-ray crystallography.

JOHN MOTTEM/APF/GETTY IMAGES
BOB STRONG/REUTERS/NEWSCOM
J. A. G. BRIGGS ET AL. 2006. *STRUCTURE* **14** (1):15–20
REPRINTED BY PERMISSION FROM SPRINGER NATURE: E. M. CAMPBELL AND T. J. HOPE. 2015. *NAT. REV. MICROBIOL.* **13** :471–83

A surprising benefit of ART was discovered through humanitarian treatment programs in Africa, such as the President's Emergency Plan for AIDS Relief (PEPFAR), initiated in 2003 by President George W. Bush and continued by subsequent presidential administrations. The PEPFAR program showed that, in communities that deliver ART to all members, regardless of infection status, the virus production decreases so much that transmission declines. Thus, while HIV cannot be eliminated from infected individuals, new infections could be near zero.

Despite available therapies, today HIV still infects one in every 200 adults worldwide, equally among women and men. HIV infection is subject to major health disparities, even in the United States, where new infections disproportionately impact Black and Latino people. In developing countries, the drugs may be expensive or unavailable for the regular continuing treatments required.

Research has long addressed the aim of an HIV vaccine, but without success. The reasons are complex:

- **High mutation rate.** The mutation rate of HIV is among the highest known for any virus. Within one patient, the virus evolves into a quasispecies (discussed later) whose different strains attack different organs and predominate at different stages of the disease.
- **Integrated HIV genome hides within cells.** The maintenance of integrated lentiviral genomes involves a greater number of regulator proteins than that required for the "simple" retroviruses.
- **HIV disables the immune system.** Because HIV infects cells of the immune system, it has various molecular mechanisms that impair both innate and active immune responses.

HIV Structure and Genome

The structure of HIV as visualized by transmission electron microscopy (TEM) consists of an electron-dense core particle (or capsid) surrounded by a phospholipid bi-layer envelope (**Fig. 11.22C**). The conical core is composed of capsid (CA) protein subunits whose arrangement is partly icosahedral (**Fig. 11.22D**). The membrane

around the core contains spike proteins , which join the membrane to the matrix, as in influenza virus. The envelope forms around the core from host cell membrane, when a progeny virion is budding out.

HIV core. The core contains two distinct single-stranded copies of the RNA genome (**Fig. 11.23A**). Unlike influenza segments, each of the two RNAs contains a complete "map" of HIV genes. However, the two RNAs of an HIV virion can have slightly different alleles arising from distinct replication events. Thus, the HIV virion is genetically "diploid." A nonfunctional mutant gene on one genome may be complemented by a functional gene on the other.

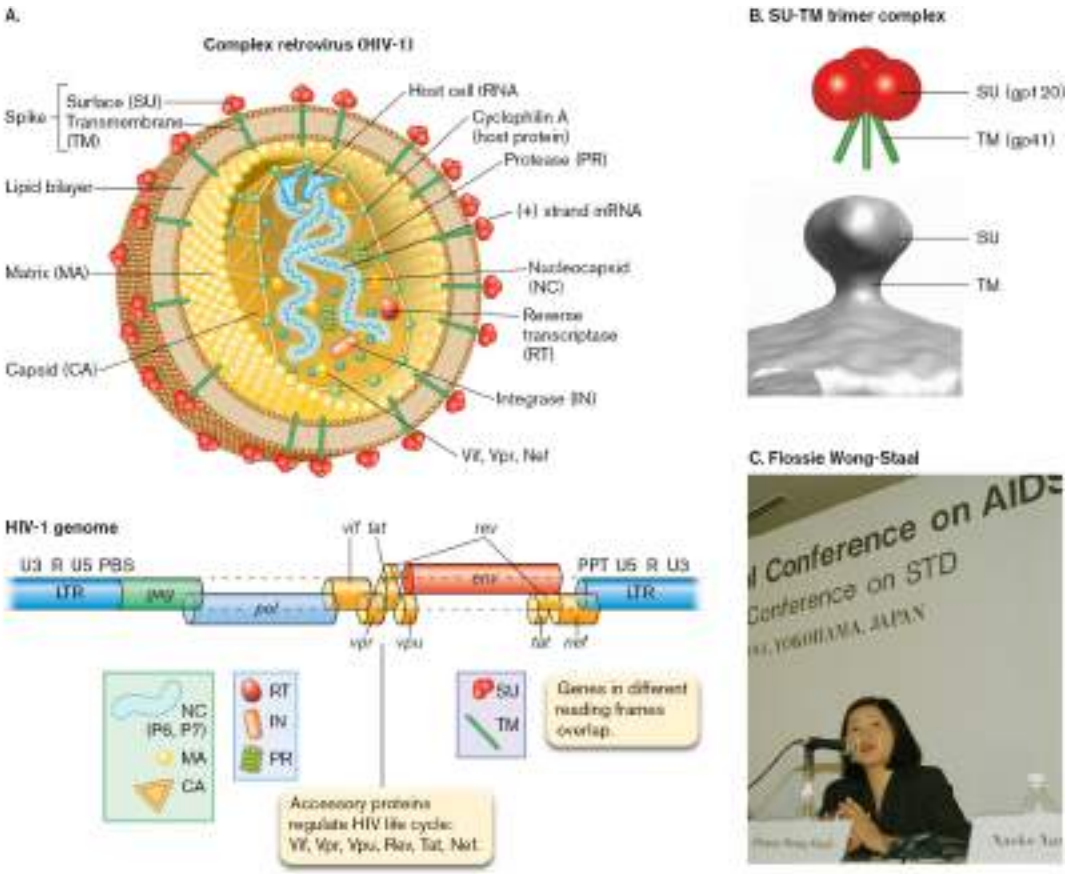


FIGURE 11.23 ■ HIV-1 structure and genome. A. Internal structure of the HIV-1 virion (top), color-coded to match the genome (bottom). In the genome sequence, the staggered levels indicate three different reading frames. LTR = long terminal repeat. **B.** Envelope spike complex (model based on cryo-EM

tomography). **C.** Flossie Wong-Staal, pioneering AIDS researcher, was the first to clone the HIV genome.

Source: Part A modified from J. A. G. Briggs et al. 2006. *Structure* **14** :15–20; part B, from P. Zhu. 2008. *PLoS Pathog.* **11** :e1000203.

AP PHOTO/KOJI SASAHARA

Each RNA genome is coated with nucleocapsid (NC) proteins similar in function to the NPs of influenza virus. Unlike influenza virus, each RNA of HIV requires a primer for DNA synthesis: a transfer RNA (tRNA) derived from the previously infected host cell. The host-derived tRNA is packaged in place on the RNA template, ready to go. The primed and packaged RNA is contained within the core composed of CA subunits. The core also contains about 50 copies of reverse transcriptase (RT) and protease (PR), as well as a DNA integration factor (integrase; IN). Unique to type HIV-1, subunits of a host chaperone named cyclophilin A are incorporated into the structure—about one for every ten core subunits. An HIV-1 mutant that fails to incorporate cyclophilin A can attach to a host cell and insert its capsid, but the core fails to come apart, and infection is halted.

The core is surrounded by a matrix (MA subunits), which reinforces the host-derived phospholipid membrane. The membrane is pegged to the matrix by spike proteins composed of the envelope subunits TM and SU (**Fig. 11.23B**). As in influenza virus, the spike proteins play crucial roles in host attachment and entry.

HIV genome. Because HIV is a high biohazard for research, it was important to clone the genome for study outside the infective virion. The genome of HIV (**Fig. 11.23A**) was first cloned by Flossie Wong-Staal (1946–2020) at the National Institutes of Health (**Fig. 11.23C**). Born in China, Wong-Staal immigrated to the United States and then worked with Gallo on the early discoveries of HIV. She later founded the Center for AIDS Research at UC San Diego.

The HIV genome includes three main open reading frames that are found in all retroviruses: *gag* , *pol* , and *env* . The *gag* sequence encodes capsid, nucleocapsid, and matrix proteins; *pol* encodes reverse transcriptase, integrase, and protease; and *env* encodes

envelope proteins. The *gag* and *pol* sequences overlap, but because they are translated in different reading frames (different ways to align the triplet code), the ribosome expresses each independently of the other. During infection, each reading frame is transcribed and translated as a polyprotein; then, at subsequent stages, each polyprotein is cleaved by proteases to form the mature products.

In HIV-1, the *gag* and *pol* sequences overlap, and *env* overlaps with genes encoding accessory proteins , proteins that modify and regulate retroviral infection. Accessory proteins are unique to lentiviruses, mediating host response and maintaining the long-term slow progression of lentiviral disease. The accessory proteins are expressed within the infected host cell and regulate the replication cycle (**Table 11.2**). For example, Tat protein activates transcription of the viral genome. The HIV-1 genome encodes at least six accessory proteins—a greater number than in any other retrovirus. They are major targets for research and drug discovery aimed at preventing HIV proliferation.

TABLE 11.2 Accessory Proteins of HIV-1

Protein	Function	Effect of mutation
Nef	Virion component: • Internalizes and degrades CD4 receptors to avoid superinfection by more HIV virions and to lessen immune response to the infected cell. • Decreases expression of major histocompatibility	Slower progression to AIDS

TABLE 11.2	Accessory Proteins of HIV-1	
	complex (MHC) proteins that stimulate cytotoxic T cells.	
Rev	Nuclear phosphoprotein, combines with host cell proteins: • Stabilizes certain mRNAs in nucleus. • Exports mRNA out of nucleus into cytoplasm, inducing shift from latent phase to virion-producing phase.	Failure of infection
Tat	Transcription factor: • Binds trans-activation response (TAR) site on nascent RNA to activate transcription. • Associates with histone acetylases and kinases to activate transcription of integrated viral DNA.	Blocks HIV transcription
Vif	Virion component: • Tags host defense protein APOBEC3G for degradation.	Virions produced are noninfective
Vpr	Virion component:	Lower production of

TABLE 11.2	Accessory Proteins of HIV-1	
	• Transcription factor; activates HIV transcription during G_2 phase of cell cycle; arrests T-cell growth. • Imports DNA across nuclear membrane; avoids need to infect rapidly dividing cells in which mitosis dissolves the nucleus.	virions
Vpu	Membrane protein: • Degrades CD4, releasing bound spike proteins. • Promotes virion assembly and release from cell-surface tetherins.	Early death of host cell; lower production of virions

Origin and evolution of HIV. Where did HIV come from? The origin of HIV has been traced back to the early twentieth century on the basis of genome sequence comparison with related viruses infecting other primates, called simian immunodeficiency viruses (SIVs; **Fig. 11.24A**). Sequence comparison of different strains of HIV and SIV reveals that an immunodeficiency virus actually entered the human population more than once, from SIV strains derived from related primates in Africa. It is thought that human consumption of primates for meat may have introduced SIV strains that then adapted to human infection. Today, the vast majority of HIV-infected patients show the

HIV-1 strain M, but some people have been infected by HIV-2, which derived independently from another SIV strain.

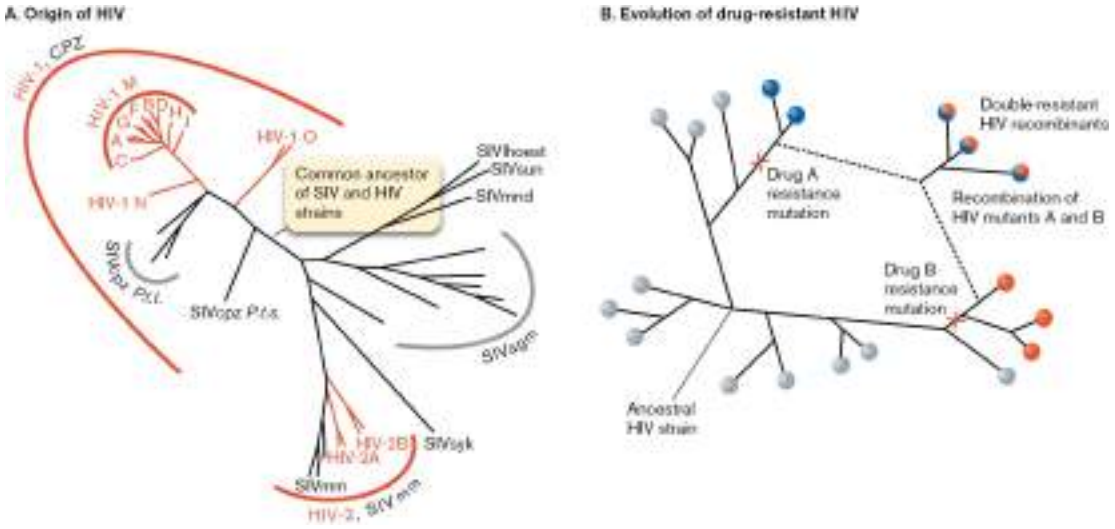


FIGURE 11.24 ■ Origin and evolution of HIV. A. Strains of HIV and SIV arose independently multiple times over several decades from a common origin in monkeys. *P.t.s. = Pan troglodytes schweinfurthii* ; *P.t.t. = Pan troglodytes troglodytes* . **B.** When HIV infects a patient, different drugs may select strains with different resistance mutations. The mutant strains may then recombine to generate a double-resistant strain.

HIV is the most rapidly evolving pathogen known; its replication generates about one mutation per progeny virion. For this reason, physicians always prescribe a combination of antiretroviral drugs, with different molecular targets. The hope is that if any one mutation confers resistance to one drug, the mutant virus will still be blocked by another. This strategy of drug combination and continual testing for resistance enables many treated HIV carriers to remain free of AIDS for decades. But in some cases, recombination of different mutants can generate strains resistant to multiple antiviral agents (**Fig. 11.24B**).

Within a single infected patient, the high mutation rates generate multiple virus strains with differing properties of replication, tissue tropism, and resistance to antibiotics. This dynamic population of diverse mutant strains is called a quasispecies (**Fig. 11.25**). The

quasispecies forms when an infective virion commences replication with rapid mutation and trait diversification. Many of the progeny virions have sequences and tropisms so different from those of their ancestor that, in isolation, they would be classified as different species. Different clonal variants of the virus colonize different tissues and organs. Virus types within a quasispecies may interact cooperatively on a functional level, by serving complementary roles in the disease state, and thus collectively define the traits of the viral population. But what happens when an infected "donor" introduces HIV to a recipient? Within the recipient, only one HIV type proliferates —close to the original type that generated the quasispecies. After this acute infection, the HIV population again diversifies into the quasispecies. It's as if a relatively narrow range of genotype carries the HIV "germ line," whereas the mutant types sustain the infected state and maintain immunosuppression.

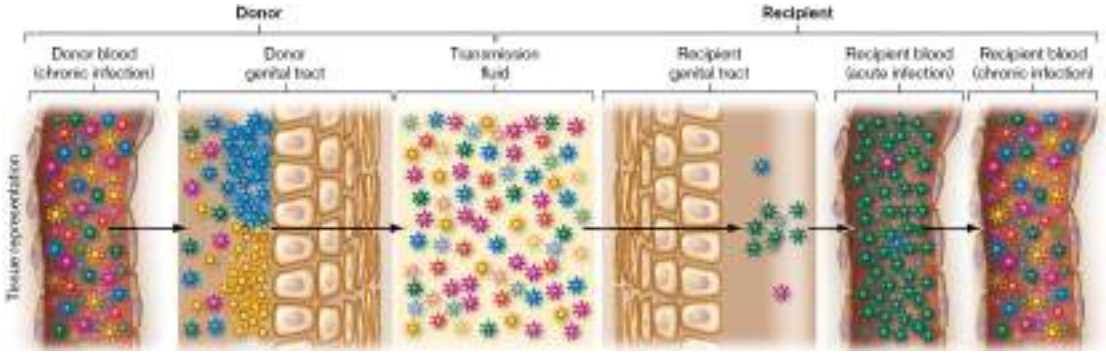


FIGURE 11.25 ■ Quasispecies development. HIV rapidly generates mutant progeny (different colors) in a chronically infected "donor." The variants may colonize different tissues and organs. Only one of these types (green) is optimized to infect the next "recipient." Within the recipient, the new virions replicate and regenerate the quasispecies.

Source: Modified from Sarah Joseph. 2015. *Nat. Rev. Microbiol.* **13** :414.

HIV Attachment and Host Cell Entry

Like other viruses, HIV needs to recognize specific receptor molecules on the surface of its target cells. The primary receptor for HIV is the CD4 surface protein on CD4 T lymphocytes (T cells). The normal function of CD4 surface proteins is to connect the T cell with an antigen-presenting cell, which activates the T cell to turn on B-cell production of antibodies (discussed in Section 24.2). Disruption of this antibody production is the main cause of the AIDS-related susceptibility to opportunistic infections. Note, however, that CD4 proteins appear on many other cell types, such as microglia (macrophage-like cells in the central nervous system) and Langerhans cells (immune cells of the epidermis). Their presence may make other cells susceptible to infection by HIV.

Spike proteins mediate membrane fusion. The binding of HIV to CD4 receptors involves the envelope spike protein SU (**Fig. 11.26**). Spike proteins are the main external proteins accessible to the host immune system.

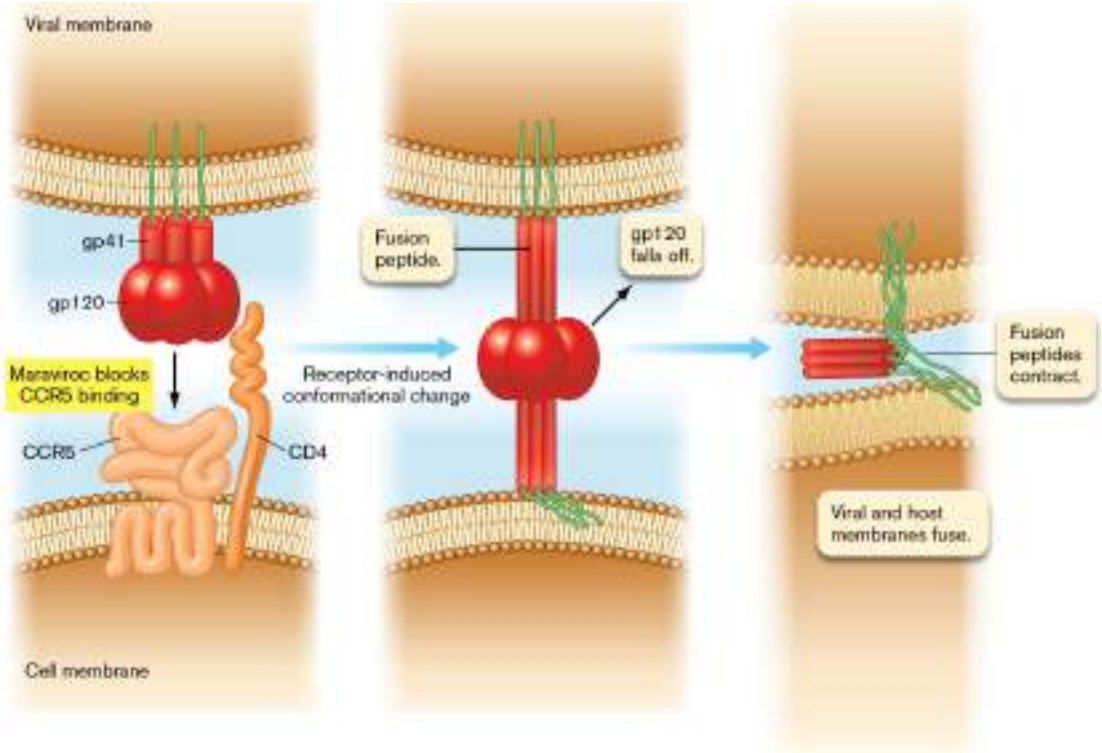


FIGURE 11.26 ■ HIV-1 attachment to host cell. The SU (gp120) subunits of the spike protein complex attach to the

receptor (CD4 cell-surface protein) and to CCR5. The fusion peptides contract, pulling the membranes together.

HIV attachment to the cell membrane requires a fusion peptide rearrangement similar to that of influenza virus, except that it takes place at the cell surface (**Fig. 11.26**). When SU (gp120) binds to CD4, the spike transmembrane component TM (gp41) unfolds and extends its fusion peptide into the host cell membrane. In addition, SU binds to secondary receptors in the membrane called chemokine receptors (CCRs) , such as the macrophage receptor CCR5. Chemokines are signaling molecules for the immune system, but their receptor proteins can bind viruses that evolve to take advantage of them. After the spike protein SU binds receptors and the TM fusion peptide inserts, the HIV-1 envelope fuses with the plasma membrane. A CCR such as CCR5 is also called a coreceptor , a protein acting with CD4 to bind the HIV spike proteins.

The requirement for CCR attachment varies among different types of HIV. Some chemokine receptors are found on neurons, and their involvement in HIV infection may mediate the neurological disorders seen in AIDS. Furthermore, the predominance of different viral envelope types with different receptor preferences varies over the course of infection. As HIV evolves a quasispecies, the virions target the CCR5 receptor early in infection, whereas later-evolved virions target a different host surface protein, called CXCR4. These "X4" virions can infect early-stage T cells that do not yet carry CCR5; thus, the AIDS disease accelerates (Section 26.1). The X4 virions are less infective when transmitted, so most new infections start with a CCR5 strain.

An exciting discovery was that individuals who lack the CCR5 protein because of a genetic defect show a high degree of resistance to HIV infection. This finding prompted the Pfizer company to develop an antiviral blocker of CCR5, maraviroc (**Fig. 11.26**), which is now used for therapy.

After HIV binds to membrane receptors, how does its genome enter the cell? The HIV envelope fuses with the cell membrane, enabling the HIV core to enter the cytoplasm directly. This HIV entry

mechanism differs from that of influenza virus, in which endocytosis and lysosome fusion are required to open the capsid and release the genome into the cell. The HIV core (composed of CA and the host-derived cyclophilin A) dissolves, releasing the two RNA genomes, along with associated viral enzymes, into the cytoplasm.

Thought Question

11.7 How do attachment and entry of HIV resemble attachment and entry of influenza virus? How do attachment and entry differ between these two viruses?

The two RNA genomes each possess a 5′ "cap" and a 3′ poly-A "tail" that enable them to mimic host nuclear mRNA. Each RNA is hybridized to a tRNA that serves as a primer for DNA synthesis. The primer is a lysine-specific tRNA from the previous infected cell. The primer might be expected to hybridize at the 3′ end of the template, where its 3′ OH "points" toward the opposite end, positioned to synthesize all the way down. Surprisingly, however, the 3′ OH end of the tRNA actually binds near the 5′ end, where initially it can generate only a brief sequence. These early sequences bind key regulatory factors for transcription and for DNA insertion into the host genome (discussed next).

Reverse Transcriptase Copies RNA to DNA

A retrovirus, unlike other RNA viruses, must integrate its entire genome into the host genome in order to replicate viral genomes and produce new progeny virions. Thus, the RNA genome needs to serve as a template to synthesize a DNA complement, but then the original RNA template must be degraded and replaced by a DNA strand for host integration. All of these processes are accomplished by reverse transcriptase (RT) , the defining enzyme of a retrovirus. Reverse

transcriptase is the source of the high error rate of retroviral replication—on average, one or two errors per copy of HIV. This high error rate generates the quasispecies of different strains within an HIV-infected person (**Fig. 11.27**).

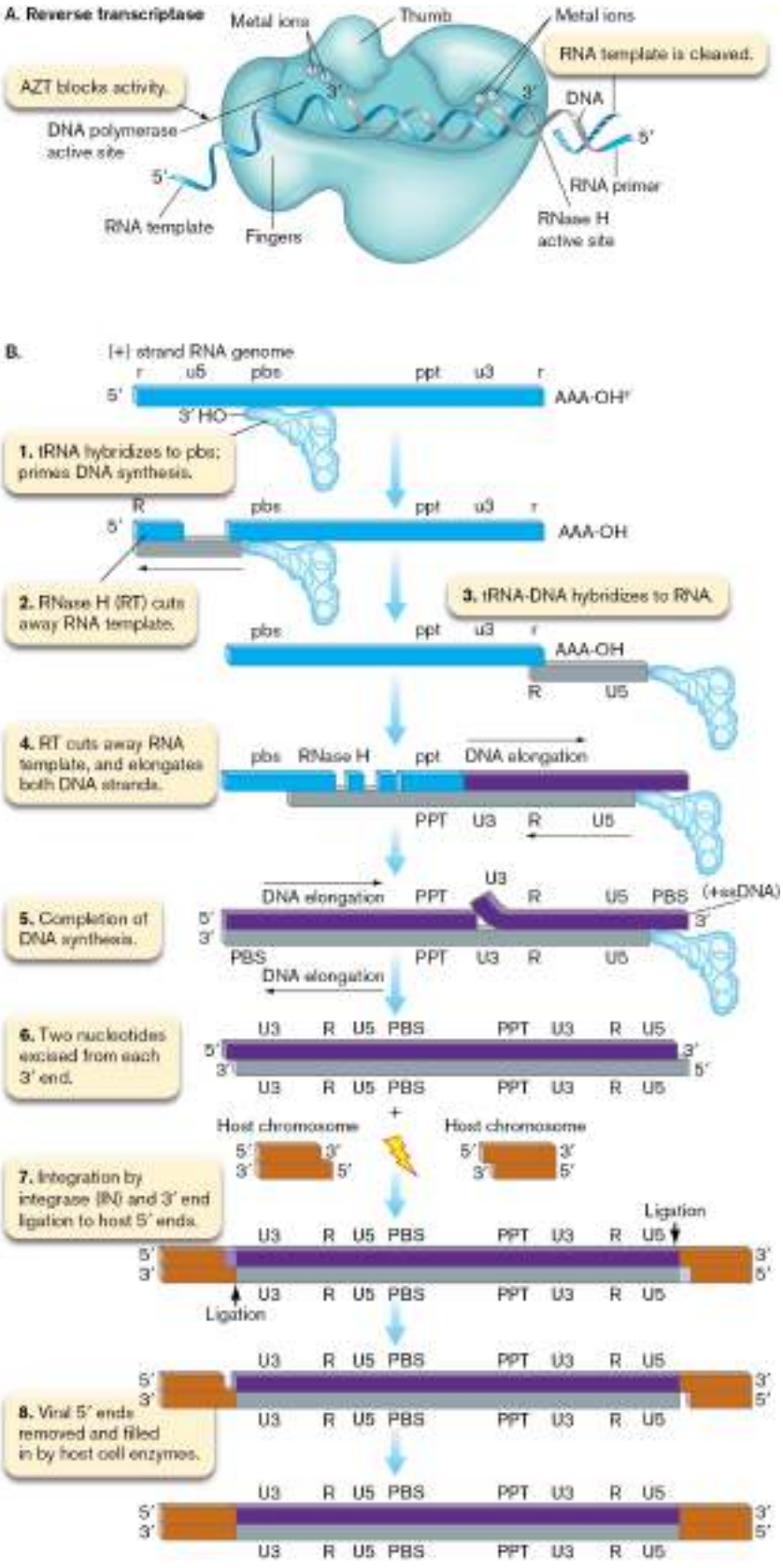

A. Reverse transcriptase
Metal ions
Thumb
Metal ions
RNA template is cleaved.
AZT blocks activity.
DNA polymerase active site
DNA
RNA primer
RNA template
Fingers
RNase H active site
B.
(+) strand RNA genome
r u5 pbs ppt u3 r
AAA-OH
1. tRNA hybridizes to pbs; primes DNA synthesis.
2. RNase H (RT) cuts away RNA template.
3. tRNA-DNA hybridizes to RNA.
4. RT cuts away RNA template, and elongates both DNA strands.
RNase H
DNA elongation
5. Completion of DNA synthesis.
(+)ssDNA
6. Two nucleotides excised from each 3′ end.
Host chromosome
7. Integration by integrase (IN) and 3′ end ligation to host 5′ ends.
Ligation
8. Viral 5′ ends removed and filled in by host cell enzymes.

FIGURE 11.27 ■ Reverse transcription of the HIV genome and integration into host DNA. A. The RNA template with its short RNA primer is threaded through the RT complex between the "thumb" and "fingers"—a configuration typical of other RNA polymerases. The RT complex adds successive dNTPs, as in regular DNA synthesis. As DNA elongates, the RNA template is cleaved from behind by the RNase H site of reverse transcriptase (RT). **B.** The tRNA primer initiates a short sequence of DNA complementary to the 5′ end of the HIV chromosome. The corresponding template is then degraded, and the DNA-primer complex is transferred to the opposite end, where it can complete synthesis of the genome. Remaining RNA is cleaved by RT and replaced by DNA. The double-stranded DNA circularizes at the U3-R-U5 sequence and then integrates somewhere in the host genome. ssDNA = single-stranded DNA.

Reverse transcriptase is the target of the first clinically useful drug to treat HIV infection, the nucleotide analog azidothymidine (AZT) . AZT is incorporated into the growing DNA chain in place of a thymidine, but because its 3′ OH is replaced by an azido group ($-N_3$), no further nucleotides can be added.

The reverse transcriptase complex actually possesses three different activities:

- **DNA synthesis from the RNA template.** Synthesis of DNA is first primed by the host tRNA, which was hybridized to the chromosome within the virion.
- **RNA degradation.** After DNA synthesis, the template RNA is gradually removed through an RNase H activity of the RT complex. Removal of RNA enables replacement of the entire original RNA template by DNA.
- **DNA-dependent DNA synthesis.** To make the DNA complementary strand replacing the RNA, the RT needs to use the newly made DNA as its template. Thus, RT has the rare ability to use either DNA or RNA as a template.

As shown in **Figure 11.27A** , the RNA template with its short RNA primer is threaded through the RT complex between the "thumb" and "fingers"—a configuration typical of other RNA polymerases (discussed in Chapter 8). The RT complex adds successive deoxynucleotides from deoxyribonucleoside triphosphates (dNTPs) starting at the 3′ OH end of the RNA primer. As DNA elongates, however, the RNA template is cleaved from behind by the RT complex. Thus, the new DNA actually replaces the preexisting RNA sequence. This "destructive replication" is unique to retroviruses. The details are important, as they suggest possible targets for new antiviral drugs.

Reverse transcription of the HIV genome. Reverse transcription of the viral genome involves several unusual mechanisms (**Fig. 11.27B**). First, the host-derived tRNA primer initiates synthesis (by reverse transcriptase) of a DNA strand complementary to the RNA chromosome. DNA is elongated toward the 5′ end of the HIV chromosome, generating a short segment (**Fig. 11.27B** , step 1). The original RNA template for this short segment is then degraded by the RNase H activity of RT, leaving only the DNA extension of the tRNA primer (step 2).

The new DNA primes the second template. The original RNA template had repeated ends (labeled "r," lowercase, for RNA in **Fig. 11.27B**), and the exposed DNA copy of the 5′ end has a complementary sequence ("R," uppercase, for DNA). The "R" DNA from the second tRNA extension hybridizes to the 3′ end of the original RNA (**Fig. 11.27B** , step 3). The hybridized DNA elongates along the rest of the chromosome (step 4), up to the primer-binding site (pbs) for mRNA transcription. DNA completion is followed by degradation of the remaining RNA template, except for occasional short fragments to serve as primers, such as the polypurine tract (ppt). A complementary DNA strand is then synthesized through the PPT primer, leaving a nick at U3 (step 5).

Notably, human host cells have evolved a protein, APOBEC3G, that interferes with reverse transcription by several mechanisms. Interference mechanisms include:

- Inhibition of tRNA priming of reverse transcription

- Deamination of cytosines to uridine in the HIV DNA product of reverse transcription, thus increasing the error rate of the HIV provirus (integrated DNA copy of the viral genome)
- Interference with removal of tRNA primer from completed DNA copy, thus inhibiting integration into the host genome

APOBEC3G can be packaged into progeny virions, and thus decreases the production of infective virions in the next host. However, HIV has evolved an accessory protein, Vif (**Table 11.2**), that binds APOBEC3G and initiates its degradation.

Integration into the host genome. The final phase of genome processing requires integration into host DNA. This complex process provides excellent targets for therapeutic inhibitors such as the drug raltegravir. Surprisingly, HIV integration requires no specific end homology, and the viral DNA generally inserts at positions with low host gene expression, such as introns. For this reason, HIV integration offers a compelling tool for gene therapy vectors (discussed in the next section).

The double-stranded DNA copy of the HIV genome undergoes integration catalyzed by the viral enzyme integrase (IN), which was packaged in the original virion (**Fig. 11.23**). First the integrase excises two nucleotides from each 3′ end of the duplex (**Fig. 11.27B** , step 6). This excision generates a 5′ overhang at each end. Integrase also nicks the host genome, generating 2-bp staggered ends. The HIV DNA then inserts at the nicked host site, with ligation of viral 3′ ends to the 5′ host ends (step 7). The viral 5′ ends are removed, and the gaps are filled in by host cell repair proteins (step 8). Overall, this mechanism forms an integrated viral genome, or provirus . The provirus includes two copies of the end sequence U3-R-U5, which is called a long terminal repeat (LTR) . Proviral sequences can now be expressed, directing production of progeny virions.

An alternative to virion production is that the integrated HIV genome lies dormant, like the lambda prophage in *E. coli* . The integrated HIV genome is replicated passively within the genome of its host cell, hiding for many years with only infrequent production of virions. The few virions shed by the patient, however, can infect an

unsuspecting individual who has sexual contact with or is exposed to the blood of the patient.

Replication Cycle of HIV

The steps of HIV replication are outlined in **Figure 11.28** . The main points of viral entry and replication are typical of retroviruses. HIV, however, has an exceptionally large number of accessory proteins that govern the level of viral production and the duration of the quiescent phase, when the integrated chromosome replicates with the host cell.

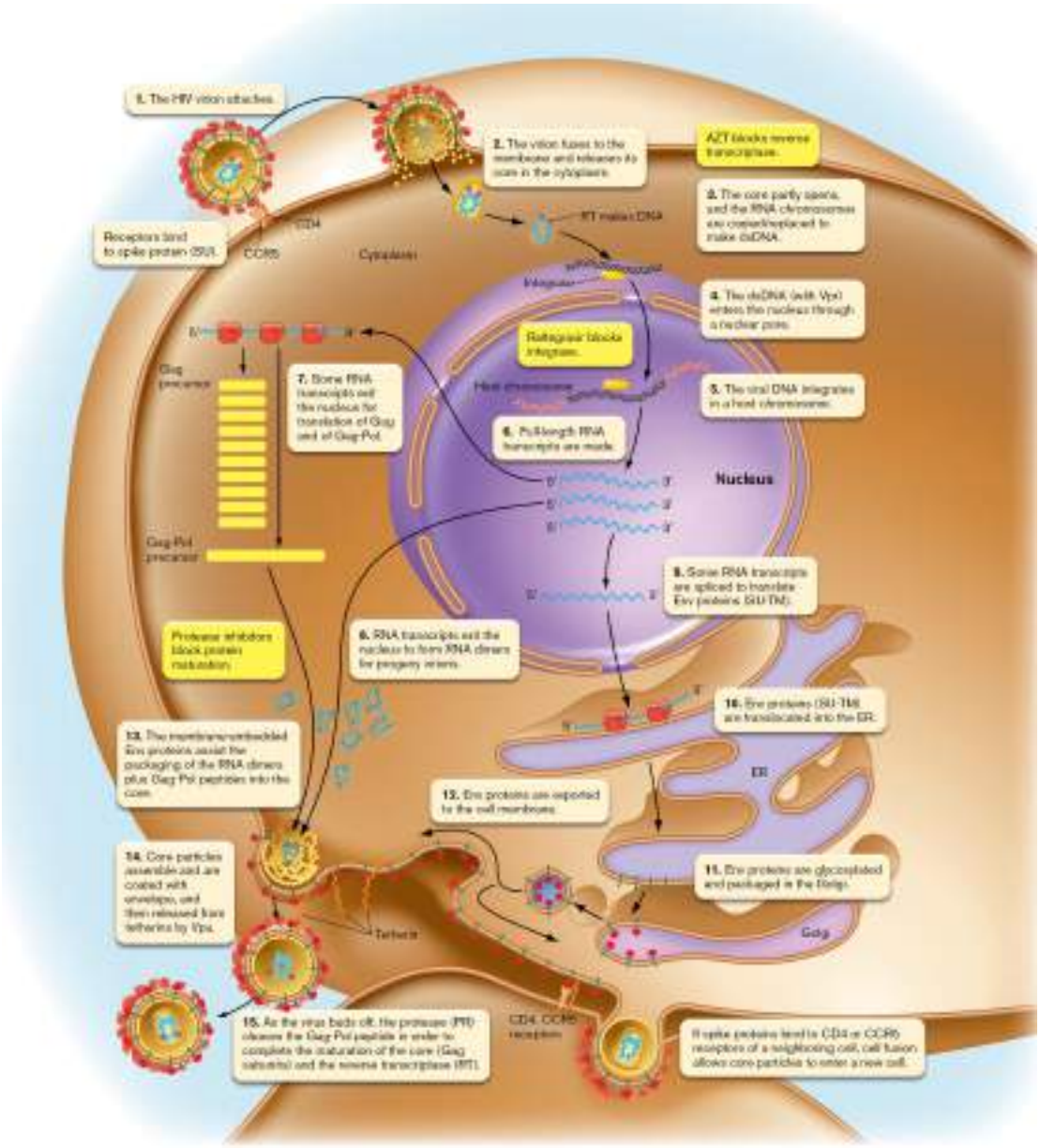


FIGURE 11.28 ■ HIV replication cycle. The HIV virion attaches its receptor and fuses with the host cell membrane, releasing its contents in the cytoplasm to undergo a replication cycle.

Synthesis of HIV mRNA and progeny genomic RNA. After the HIV virion attaches to the host receptors, its envelope fuses with the host membrane (**Fig. 11.28** , step 1). Unlike influenza virus, the HIV core enters the cytoplasm directly, without endocytosis (step 2). The

core partly uncoats, while the RNA chromosomes within are reverse-transcribed to make double-stranded DNA (step 3). The double-stranded DNA enters the nucleus through a nuclear pore (step 4)—a key step facilitated by Vpr accessory protein. Vpr enables infection of nondividing cells, which only lentiviruses can do; other retroviruses, such as those causing lymphoma, must infect dividing cells, in which the nuclear membrane dissolves during mitosis.

Upon entering the nucleus, the DNA copy of the HIV genome integrates its sequence as a provirus at a random position in a host chromosome (**Fig. 11.28** , step 5). Integration is catalyzed by integrase (the IN protein; **Fig. 11.23**). Integrase inhibitors such as raltegravir are an important class of anti-HIV drugs. Within the nucleus, full-length RNA transcripts are made by host RNA polymerase II, including a 5′ cap and a 3′ poly-A tail (step 6). Some of the RNAs exit the nucleus (step 7) to serve as mRNA for translation of polyproteins. Polyproteins are translated in alternative versions, such as Gag-Pol. Other full-length RNA transcripts exit the nucleus to form RNA dimers for progeny virions (step 8). Still other RNA transcripts within the nucleus are cut and spliced to complete the *env* gene sequence for translation of Env (envelope) proteins (step 9).

The Env proteins are made within the endoplasmic reticulum (**Fig. 11.28** , step 10). They pass through the Golgi for glycosylation and packaging (step 11) and are exported to the cell membrane (step 12). At the membrane, Env proteins plug into the core particle as it forms from the RNA dimers plus Gag-Pol peptides (step 13).

Virion assembly and exit. The core particles are packaged with envelope derived from host cell membrane containing Env spike proteins (**Fig. 11.28** , step 14). To escape the host cell, emerging virions require the accessory protein Vpu to bind a "tetherin," a host adhesion protein induced by interferon to cause reuptake and digestion of virions. Vpu causes proteasomal degradation of the tetherin. Emerging virions, some still tethered to the cell, are shown in **Figure 11.29A** .

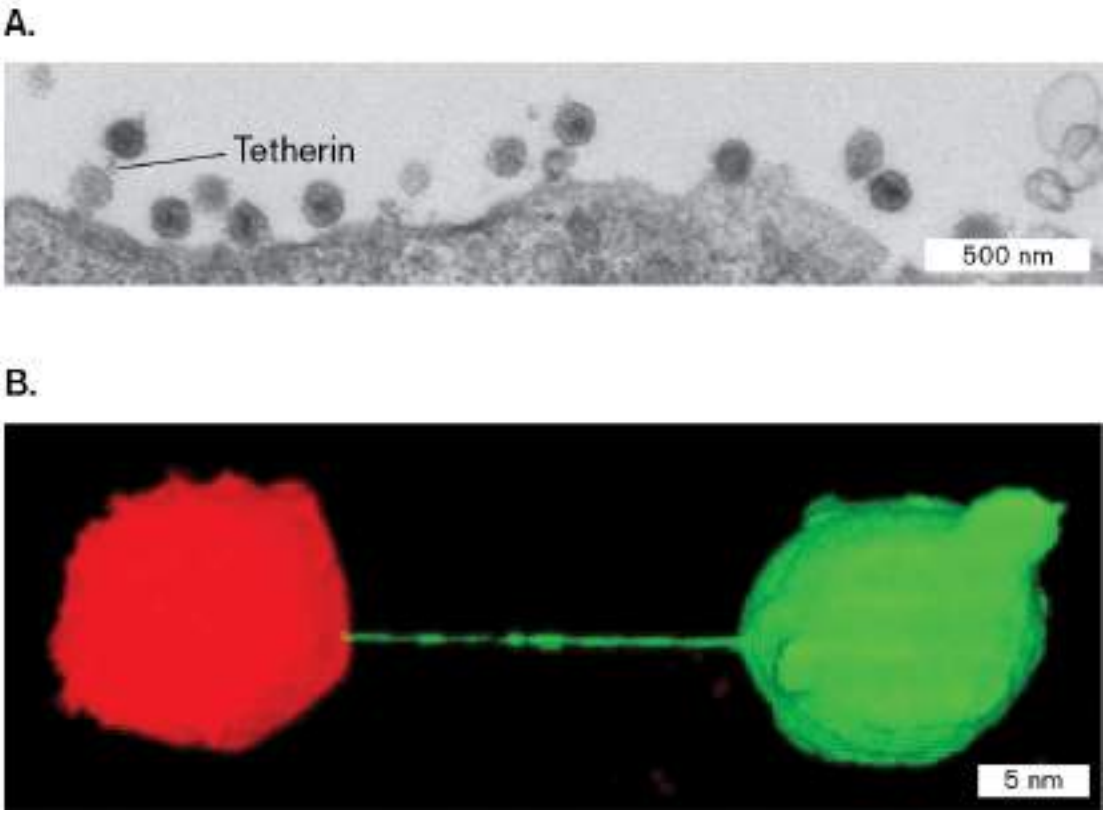


FIGURE 11.29 ■ HIV exits from an infected cell. A. As virions emerge, they remain tethered to the cell surface by host tetherins, requiring a release step mediated by accessory protein Vpu (TEM). **B.** A T cell infected with HIV can transfer virions to an uninfected cell through a nanotubular connection. The two T cells are tagged here with different fluorescent labels (red versus green).

STUART NEIL ET AL. 2008. *NATURE* **451** :425
STEFANIE SOWINSKI ET AL. 2008. *NAT. CELL BIOL.* **10** :211

As the virion buds off, the protease (PR) cleaves the Gag-Pol peptide to complete maturation of the core structure containing Gag subunits, as well as maturation of reverse transcriptase (RT; **Fig. 11.28** , step 15). The Gag subunits now form the conical core structure. Proteases that cleave Gag-Pol offer important drug targets, which have led to the development of anti-HIV drugs known as protease inhibitors .

However, HIV has alternative means of cell-to-cell transmission that avoid exposing virions to the immune system. One alternative is cell fusion, mediated by binding of Env in the membrane to CD4 receptors on a neighboring cell (**Fig. 11.28**). The two cells then fuse, and HIV core particles can enter the new cell through their fused cytoplasm. The fusion of many cells can form a giant multinucleate cell called a syncytium. Cell fusion with formation of syncytia enables HIV to infect neighboring cells without ever exiting a cell. Another means of cell-to-cell transmission is to travel through a "nanotube" connection between two T cells (**Fig. 11.29B**).

The intricate scheme in **Figure 11.28** actually omits many functions of HIV accessory proteins that enhance the virulence of HIV infection, especially their interactions with host cell components (**Table 11.2**). Mutation of genes for accessory proteins often decreases virulence; thus, these proteins are potential targets for chemotherapy. Surprisingly, even a modest decrease of HIV infectivity can have major benefits for the patient, suggesting that HIV is so crippled by its high mutation rate that the slightest interference greatly decreases production of infective virions. Thus, numerous effective antiviral agents are now known—but all have side effects, and all select for resistant strains.

To Summarize

- **Human immunodeficiency virus (HIV)** is the cause of an ongoing pandemic of acquired immunodeficiency syndrome (AIDS). Molecular biology has led to drugs that control the infection.
- **HIV is a retrovirus** whose RNA genome is reverse-transcribed into double-stranded DNA, which integrates into the DNA of the host cell. HIV evolved from simian retroviruses.
- **The HIV core** contains two different copies of its RNA genome, each bound to a primer (host tRNA) and reverse transcriptase

(RT). The core is surrounded by an envelope containing spike protein trimers.

- **HIV binds the CD4 receptor** of T lymphocytes together with the chemokine receptor CCR5. After virion-receptor binding and envelope-membrane fusion, the HIV core particle is released into the cytoplasm, where it partly uncoats.
- **Reverse transcriptase synthesizes DNA from the HIV RNA template** , primed by the tRNA. **RNA degradation** enables formation of a double-stranded DNA. Entering the nucleus, the retroviral **DNA integrates** into the host genome.
- **Retroviral mRNAs are exported to the cytoplasm** for translation. Envelope proteins are translated at the endoplasmic reticulum and exported to the cell membrane.
- **Retroviruses are assembled at the cell membrane** , where virions are released slowly, without lysis. Alternative routes of cell-to-cell transmission involve cell fusion (forming syncytia) or travel through an intercellular nanotube.
- **Accessory proteins regulate virion formation** and the latent phase, in which double-stranded DNA persists without reproduction of progeny virions.

Glossary

retrovirus
: Also called *RNA reverse-transcribing virus* . A single-stranded RNA virus that uses reverse transcriptase to generate a double-stranded DNA.

reverse transcriptase (RT)
: An enzyme that produces a double-stranded DNA molecule from a single-stranded RNA template.

human immunodeficiency virus (HIV)
: A human-specific retrovirus that causes AIDS.

acquired immunodeficiency syndrome (AIDS)

A disease caused by HIV that leads to the destruction of T cells and the inability to fight off opportunistic infections.

lentivirus

A member of a family of retroviruses with a long incubation period. An example is HIV.

endogenous retrovirus

A retroelement (genome sequence descended from a retrovirus) that contains *gag* , *env* , and *pol* genes.

core particle

A viral capsid that encloses its nucleic acid genome and is surrounded by an envelope.

spike protein

A viral glycoprotein that connects the membrane to the capsid or the matrix and may be involved in viral binding to host cell receptors.

accessory protein

A protein found in the viral capsid or tegument that is needed early in the viral life cycle.

quasispecies

A collection of isolates (usually viruses) from a common source of infection that have evolved into many different types within one host.

chemokine receptor (CCR)

A human T-cell membrane protein that binds chemokine hormones but is also used by HIV for attachment and infection.

coreceptor

A cell-surface receptor needed for viral entry along with a primary receptor.

reverse transcriptase (RT)

An enzyme that produces a double-stranded DNA molecule from a single-stranded RNA template.

azidothymidine (AZT)

A nucleotide analog that inhibits reverse transcriptase and was the first drug clinically used to fight HIV infections.

provirus

A viral genome that is integrated into the host cell genome.

long terminal repeat (LTR)

A repeated nucleic acid sequence at the 5′ and 3′ ends of a provirus.

protease inhibitor

A molecule that inhibits a protease enzyme; some are used as anti-HIV drugs to block the virally encoded protease needed to complete HIV assembly.

11.4 Endogenous Retroviruses and Lentiviral Gene Therapy

Suppose an integrated HIV genome mutated and lost the ability to produce progeny. What would happen to its genome? The integrated genome would be "trapped" within a cell, an endogenous retrovirus. If the cell entered the host germ line, over many host generations in its host the retroviral sequence would inevitably accumulate more mutations. It could even provide the material for evolution of a new trait. Such endogenous retroviruses inspired the idea that researchers could intentionally manipulate a retrovirus to impart a useful trait and use it for gene therapy.

Retroelements in the Human Genome

The human genome is riddled with remains of retroviral genomes in various states of decay. These decaying genomes are collectively known as retroelements (**Fig. 11.30**). Endogenous retroviruses (in humans, HERVs) are retroelements that retain all the genomic elements of a retrovirus, including *gag* , *env* , and *pol* genes. Other endogenous elements retain part of the retroviral genome but have lost essential sequences through reductive evolution (discussed in Chapter 17).

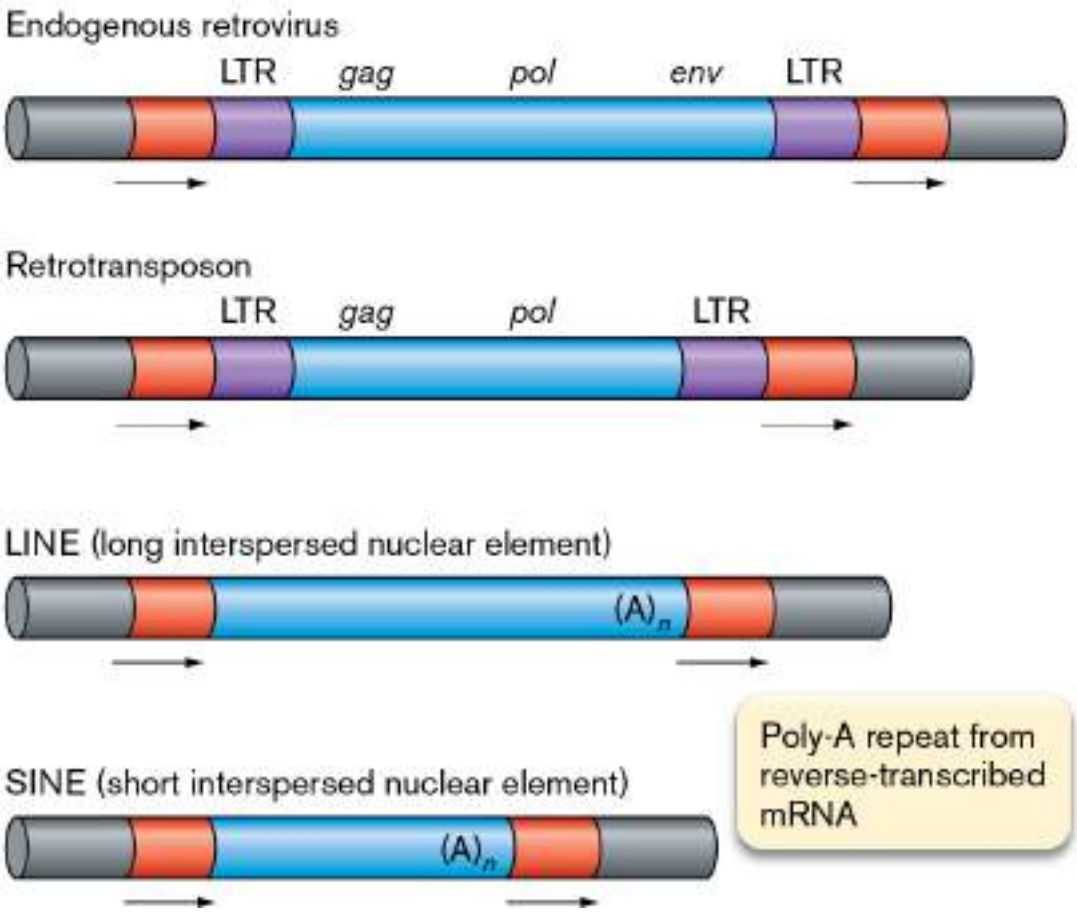


FIGURE 11.30 ■ Retroelements in the human genome. Endogenous retroviruses and other retroelements in the human genome may arise from progressive degeneration of ancestral retroviruses or they may be progenitors of new retroviruses.

Retrotransposons retain only partial retroviral elements but may maintain a reverse transcriptase to copy themselves into other genome locations. An example of a retrotransposon is the well-known Alu sequence, a short sequence found in about a million copies in the human genome. In some cases, a retrotransposon such as Alu can interrupt a key human gene, leading to a genetic defect such as a defective lipoprotein receptor associated with abnormally high cholesterol level and heart failure. Still other retroelements, known as LINEs (long interspersed nuclear elements) and SINEs (short interspersed nuclear elements), show more vestigial remnants of retroviral genomes. Amazingly, retroelements and transposons

appear to have generated about half the sequence of the human genome.

Remarkably, many human endogenous retroviruses (HERVs) and other mammalian endogenous retroviruses (ERVs) express viral proteins that have evolved into required parts of the normal host function. For example, the human embryo requires ERVW-1 to express a protein called syncytin, which enables fusion of placental and maternal cells. The gene encoding syncytin is derived from the retroviral gene for its envelope protein; in the original retrovirus, this protein causes fusion of infected host cells.

In another case, cells expressing a normal endogenous retrovirus form particles that appear to be actual virions. HERV-K proteins are expressed by the embryonic blastocyst (**Fig. 11.31**). At the blastocyst stage, cells express proteins of the endogenous retrovirus HERV-K and form virus-like particles (labeled red in **Figure 11.31** by HERV-K capsid antibody stain). The virus particle formation is not random, but highly regulated; HERV-K particles arise when cell nuclei express an embryonic regulator protein (labeled green). The function of these particles is unknown, but viral proteins might protect the embryo from infection by exogenous viruses.


GFP

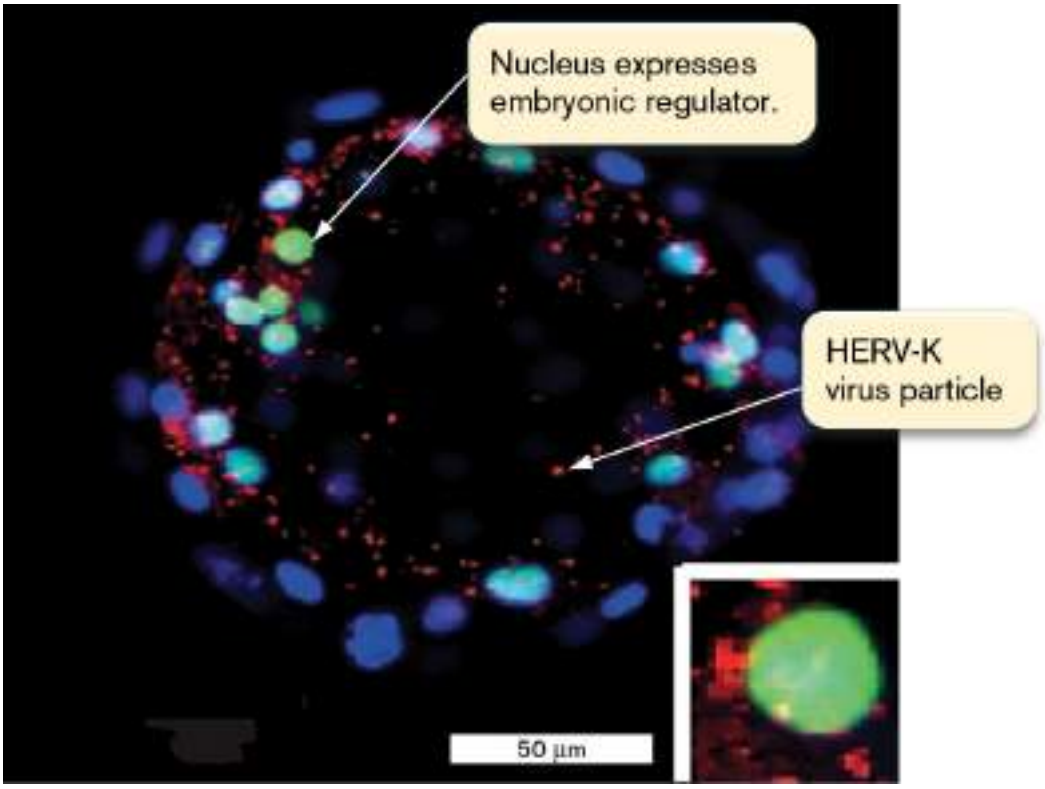


FIGURE 11.31 ■ Human embryo makes retroviral particles. Human embryos at the blastocyst stage express proteins of the endogenous retrovirus HERV-K, including virus-like particles labeled by HERV-K capsid antibody stain (red). The blastocyst nuclei are labeled with DAPI fluorophore (blue). HERV-K virus-like particles arise when cell nuclei express an embryonic regulator protein (labeled green). **Left inset:** Joanna Wysocka, a stem cell biologist at Stanford University, studies the contribution of ERVs to the human genome. **Lower right inset:** Enlarged view of embryonic regulator protein.

COURTESY OF JOANNA WYSOCKA
E. GROW AND J. WYSOCKA ET AL. 2015. *NATURE* **522** :221

On the downside, some HERV sequences may have negative consequences. Some evidence associates HERV-K expression of endogenous virus particles with neurodegenerative diseases such as amyotrophic lateral sclerosis (ALS). ALS patients show reverse transcriptase activity and elevated expression of a HERV locus (HML6

on chromosome 3). In one case, an ALS patient who was undergoing antiretroviral therapy for HIV experienced disappearance of his ALS symptoms, implying that the antiretroviral therapy had eliminated endogenous retroviruses. This possibility has stimulated some exciting research on neurodegenerative diseases.

Viral Gene Therapy

For millions of years, viruses have interacted with the genomes of humans and our prehuman ancestors—often contributing valuable genes whose products enhance our fitness. Could the genetic properties of such viruses be modified by our technology—to engineer therapy for medicine? Increasingly, the amazing answer is yes. Viruses are engineered as gene transfer vectors to deliver genes into our genome. A gene transfer vector is a DNA sequence that can express a recombinant gene within an animal or plant cell, either from a plasmid or from a sequence integrated into the host genome.

Gene transfer vectors are constructed from viruses whose replication cycles establish a viral genome within the host nucleus. Various kinds of viral vectors have produced exciting therapeutic results (**Table 11.3**). The first viral vectors were made from double-stranded DNA viruses such as adenoviruses, which cause mild respiratory illness. An adenoviral genome enters the cell nucleus, where it circularizes and replicates separately from the host chromosomes, in a cycle similar to that of herpesviruses (see Section 11.5). Thus, adenoviral vectors avoid the long-term risks of inserting DNA permanently into the genome of the host cell. A promising example in 2017 was the use of an adenoviral vector to express protein SMN in infants with spinal muscular atrophy type 1.

TABLE 11.3	**Gene Therapy with Viruses: Examples**	
Disease treated	**Vector type**	**Reference**
Spinal muscular atrophy type 1	Adeno-associated virus type 9	J. R. Mendel et al. 2017. *N. Engl. J. Med.* **377** :1713
Adenosine deaminase severe combined immunodeficiency (ADA-SCID)	Gammaretroviral vector (Strimvelis)	A. Aiuti et al. 2017. *EMBO Mol. Med.* **9:** 737
Acute lymphoblastic leukemia (ALL)	HIV-derived lentivector for chimeric antigen receptor (CAR) T-cell therapy	S. L. Maude et al. 2014. *N. Engl. J. Med.* **371** :1507
Sickle-cell disease	HIV-derived lentiviral vector (LentiGlobin BB305)	J.-A. Ribeil et al. 2017. *N. Engl. J. Med.* **376** :848

A disadvantage of adenoviral vectors is that the adenoviral genes are eventually lost from the recipient, and thus the treatment generally must be repeated. Repeated exposure to the vector eventually stimulates an immune response that destroys it. Other challenges for adenoviral vectors involve the cellular trafficking of the virus and its gene product.

Lentiviral Gene Therapy

Another exciting class of gene transfer vectors derives from the lentivirus HIV. Vectors in this class are called lentiviral vectors, or lentivectors . Lentivectors integrate genes into a host chromosome, providing longer-lasting therapy (**Fig. 11.32**). Vectors derived from HIV are particularly useful for their ability to transfer genes into nonmitotic cells. And, surprisingly, the HIV integrase has a preference for avoiding integration into highly transcribed host genes; thus, lentivectors are less likely to interfere with host function. The basis of this property of integrase remains unknown, but it is highly useful for gene therapy.

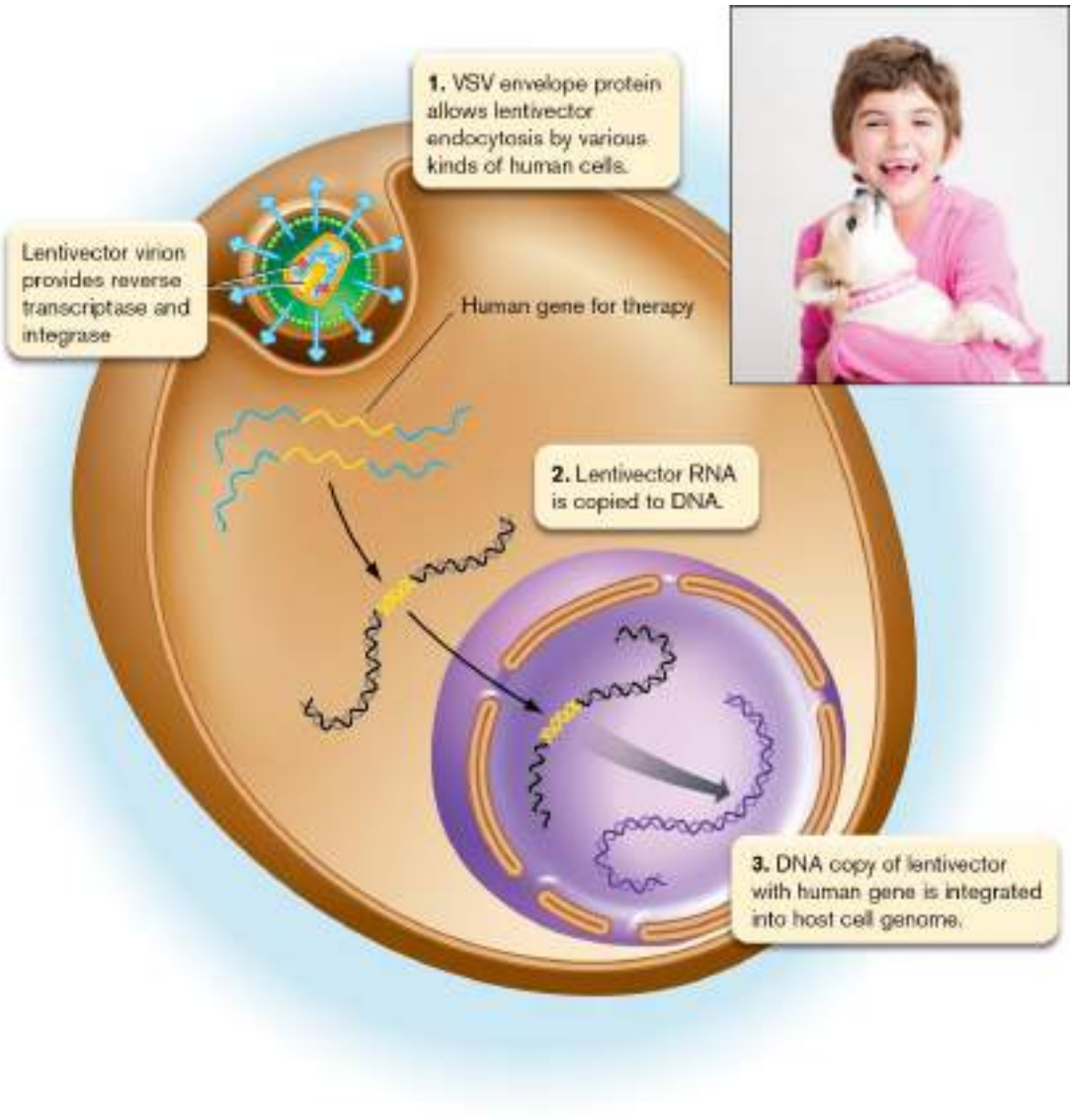


FIGURE 11.32 ■ Lentiviral gene therapy. A lentiviral vector (lentivector) derived from HIV consists of virions coated with a VSV (vesicular stomatitis virus) envelope protein that enables uptake by various kinds of cells. The engineered viral genome lacks disease-causing genes but possesses a transgene needed by the patient. The lentivector RNA with the transgene is copied into DNA and integrated into a host chromosome. **Inset:** Emily Whitehead was the first child to be considered cured of an illness (B-cell leukemia) by a lentiviral vector.

EMILY WHITEHEAD FOUNDATION

For therapeutic use, the native lentiviral genome requires extensive modification. The lentivector is engineered to remove viral genes that cause disease and to express an altered envelope protein from a different virus, such as vesicular stomatitis virus (VSV; **Fig. 11.32**). The VSV envelope protein increases the viral host range and tropism, allowing treatment of a wide range of tissues. The vector contains a human transgene that becomes integrated into the host genome.

Could lentiviral integration be dangerous for the patient? Lentiviral integration into host DNA poses the danger of activating an adjacent proto-oncogene; that is, a human gene that causes cancer when expressed at the wrong time. To avoid this problem, lentivectors are modified so that their promoters can activate only the gene of choice, not an adjacent cancer gene. Furthermore, lentiviral vectors offer a way to integrate DNA into nondividing cells of differentiated tissues such as brain neurons. For example, in 2009 a lentivector derived from HIV halted progression of a fatal brain disease, adrenoleukodystrophy (ALD), in two young boys. The lentivector inserted a gene, replacing a defective gene in each boy's blood stem cells.

In a different case, in 2012, the first leukemia patients were successfully treated by lentiviral gene therapy. The HIV-derived vector reprogrammed the patients' own T cells to attack cancerous B cells. A 7-year-old girl, Emily Whitehead (**Fig. 11.32,** inset), was the

first child to be considered fully "cured" of disease by a lentiviral vector. The commercial development of this lentiviral therapy, known as CAR-T therapy, is described in Chapter 16 .

CAR-T therapy is now approved by the FDA for B-cell malignancy. Lentivectors are also being explored for use in other cancers, such as pancreatic cancer and non-Hodgkin's lymphoma, and in experimental trials for genetic diseases including beta-thalassemia, X-linked adrenoleukodystrophy, and Wiskott-Aldrich syndrome. An exciting new approach is the use of lentivectors to deliver gene editing, as in guide RNAs for CRISPR-Cas9 (eResearch Activity 11).

How a Lentivector Works

To construct a safe and effective vector, the HIV genome is modified extensively. In the example shown in **Figure 11.33** , accessory genes *vpr* , *vpu* , *nef* , and *vif* , which encode HIV virulence factors for disease, were removed. Other protein-encoding genes necessary for virion production were put into DNA helper plasmids, to be provided only in tissue culture for vector production. These genes provide the capsid monomer (*gag*), reverse transcriptase (*pol*), envelope glycoprotein (*env*), and a regulator of mRNA export from the nucleus (*rev*). The HIV *env* gene is replaced by an *env* gene from another virus, vesicular stomatitis virus (VSV), as described earlier. The broad tropism of VSV envelope protein enables the lentivector to infect a broad range of host cell types. Each helper plasmid drives its gene expression from a well-studied promoter of another virus, such as cytomegalovirus (CMV) or respiratory syncytial virus (RSV).

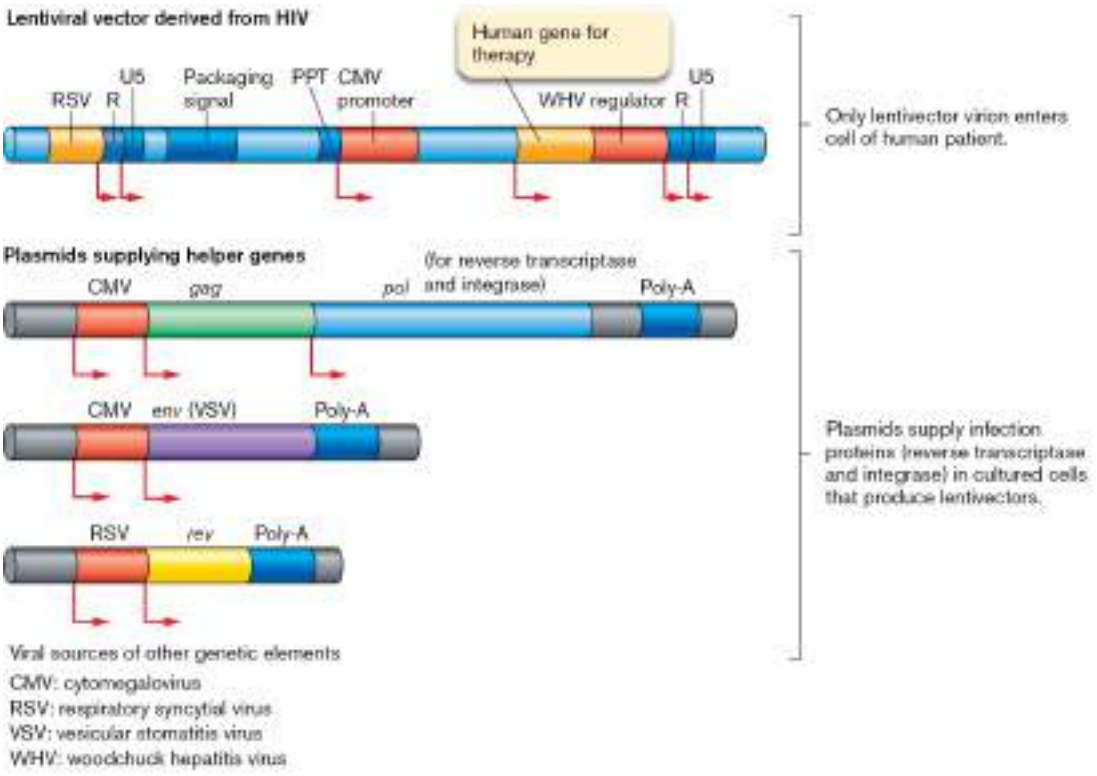


FIGURE 11.33 ■ Lentivector with helper plasmids. The lentivector genome consists of an RNA sequence containing HIV signal elements required for genomic integration (dark blue), promoter and regulator elements (red) derived from various other viruses, and the therapeutic human transgene (orange). In order to produce the virions in cell culture, essential HIV genes are provided on DNA helper plasmids. Their expression is driven by regulatory elements from other viruses (red).

Source: Modified from A. Blesch. 2004. *Methods* **33** :164.

See above for *Construction of a Gene Therapy Vector* animation

The lentivector genome retains only the LTR end sequences (R-U5) required for genome integration, a packaging signal from the start of *gag* , and the infectivity-enhancing polypurine tract (PPT). The lentivector virion (**Fig. 11.32**) packages reverse transcriptase and integrase, originally expressed by the helper plasmids in tissue culture.

To express the transgene (the gene of interest for expression in the host), a CMV promoter was inserted in the HIV-derived RNA vector. To further enhance transgene expression, a genetic enhancer sequence was added from woodchuck hepatitis virus (WHV). Note that a "lentiviral" gene transfer system, in fact, includes genetic elements from a diverse set of human and animal viruses, all found in previous studies to contribute specific properties to infection. Although these genes originate from different viruses, they nevertheless function together like parts of a machine.

To produce infective virions, the HIV-derived RNA vector plus the three helper plasmids are introduced into a special tissue culture line. The vector and plasmids enter the cells by a process called transfection , in which calcium phosphate treatment promotes the uptake of nucleic acids across the cell membrane. The host tissue culture cells are derived from human embryonic kidney 293T cells containing a gene for a protein from simian virus 40 (SV40) that enables the replication of DNA plasmids containing an SV40 replication site. The 293T cells allow efficient expression of the viral genes on the helper plasmids, as well as a full virion production cycle in which the vector RNA is replicated and packaged into virions.

Safety of Lentivectors

The use of disease-causing viruses for human therapy raises concerns about safety. The viruses might express toxins, induce cancer, or trigger a damaging immune response. These concerns need to be weighed against the risks of the conditions they are used to treat, such as severe combined immunodeficiency (SCID), cystic fibrosis, and cancer. In general, gene therapy is approved to treat only life-threatening conditions for which alternative therapies are inadequate.

Viruses used for gene therapy are engineered extensively to decrease risks. Safety features of viral vectors include:

- **Deleting virulence genes.** Viral genes that promote disease and virion proliferation, but are not required for establishment of the viral DNA in the nucleus, are deleted from the vector

genome. To produce the vector in tissue culture, the viral proliferation genes are provided on helper plasmids.

- **Avoiding genome insertion next to oncogenes.** Vectors engineered from adenoviruses are usually designed to avoid host chromosome integration altogether. The disadvantage of avoiding integration is that the separate viral DNA is soon lost from host tissues, and the therapy requires frequent repetition. In lentivectors, molecular modification avoids the activation of proto-oncogenes.
- **Altering tissue specificity.** The tropism, or tissue specificity, can be altered by replacing the gene for the viral envelope glycoprotein (spike protein) with the envelope gene from a different virus. For example, a rabies virus glycoprotein can be used to target the vector to brain cells. The alteration of viral tissue specificity by envelope gene replacement is called "pseudotyping." Pseudotyping can be used either to narrow the host range or to broaden it, depending on the needs of the vector.
- **Avoiding germ-line infection.** Current medical standards prohibit alteration of the germ line, the egg and sperm cells that transmit genes to the next generation. Because the long-term risks of gene therapy are unknown, only somatic gene therapy (gene insertion into somatic, or body, cells) is permitted.

Thought Question

11.8 What do you think are the arguments for or against lentivectors editing the germ line?

To Summarize

- **Ancient retroviral sequences persist within animal genomes** , including the human genome. Endogenous retroviruses (ERVs) retain the four main retroviral genes: *env* , *gag* , *pol* , *pro* (protease). Other sequences decay by mutation and are called retroelements. Retrotransposons can transcribe themselves into new locations in the host genome.
- **Endogenous retroviruses evolve to express essential human genes.** Human endogenous retrovirus (HERV) overexpression is also associated with neuropathologies such as ALS.
- **Gene transfer vectors** are made from viruses.
- **Virulence genes are deleted from the genome of a lentivector (lentiviral vector).** Efficient promoter sequences from other viruses are inserted. The lentivector virion packages reverse transcriptase and integrase expressed by helper plasmids.
- **Adenoviral vectors circularize and replicate separately from the host genome.**
- **Lentivectors integrate within the host genome but cannot produce progeny virions.**

Glossary

endogenous retrovirus
: A retroelement (genome sequence descended from a retrovirus) that contains *gag* , *env* , and *pol* genes.

retrotransposon
: A retroelement that contains only partial retroviral sequences but may encode reverse transcriptase to allow further movement into the host genome.

gene transfer vector
: A mobile DNA engineered from a virus or plasmid, designed to insert a genetic sequence into the genome of an organism for experimental study or for medical therapy.

lentivector *or* lentiviral vector

: A gene transfer vector derived from a lentivirus such as HIV; designed to integrate genes into a host chromosome.

transgene

: A gene that has been transferred by genetic engineering techniques from one organism to another.

transfection

: In biotechnology, the transfer of DNA (usually viral) into cells.

11.5 Herpes Simplex Virus: DNA Virus

Many important viruses of humans and other animals contain genomes of double-stranded DNA (**Table 11.4**). DNA viruses include the causative agents of well-known diseases such as smallpox, chickenpox, and infectious mononucleosis ("mono"). In 2022, the poxvirus monkeypox emerged in a global epidemic of disease spread by intimate contact.

TABLE 11.4	**DNA Viruses of Animals (Examples)**		
Virus	**DNA replication**	**Disease(s)**	**Host(s)**
Adenoviruses (many strains)	Viral DNA polymerase, single-strand binding protein, and protein primer	Enteritis or respiratory diseases	Humans, other mammals, birds
Papovavirus (simian virus 40; SV40)	Cellular DNA polymerases	Asymptomatic	Monkeys
Herpesviruses			
Herpes simplex virus 1 and 2	All viral componen	Epithelial and genital	Humans

TABLE 11.4	DNA Viruses of Animals (Examples)		
	ts (DNA polymerase, primase, etc.)	lesions, latency in neurons	
Varicella-zoster virus	Viral components	Chickenpox, shingles	Humans
Epstein-Barr virus (EBV)	All cell components (DNA polymerase, etc.)	Infectious mononucleosis, Hodgkin's lymphoma	Humans
Other strains	Varies	Epithelial lesions, cancer	Monkeys, cattle, horses
Papillomaviruses	Viral DNA helicase; cellular polymerase		
Human papillomaviruses (many strains)		Genital warts, cervical and penile cancer, skin warts	Humans
Other papillomaviruses		Warts, cancer	Rabbits, cattle, sheep

TABLE 11.4	DNA Viruses of Animals (Examples)		
Poxviruses	All viral components		
Variola major virus		Smallpox	Humans
Vaccinia virus		Cowpox	Cattle, humans
Monkeypox (mpox) virus		Monkeypox (mpox)	Rodents, monkeys, humans

Most DNA viruses are considerably larger than RNA viruses and encode a wider range of viral enzymes; for example, the vaccinia genome encodes nearly 200 different proteins. The complexity of viruses such as vaccinia and herpes approaches that of small cells.

An important class of DNA viruses infecting humans is that of herpesviruses. Herpesviruses of many kinds have been associated with humans and our ape ancestors for hundreds of millions of years (see Chapter 6 , **Figure 6.20**). Different herpesviruses cause diseases ranging from chickenpox (varicella-zoster virus) to birth defects (cytomegalovirus). Herpesviruses such as Epstein-Barr virus infect nearly all humans by adulthood. Certain strains may assist development of our immune system. Today, in a remarkable twist, we have engineered a "tumor-eating herpes" to treat metastatic tumors.

Herpes Simplex Virus Infects the Oral or Genital Mucosa

An important example of a DNA virus is herpes simplex virus (HSV). Strains HSV-1 and HSV-2 cause one of the most common infections in the United States. Approximately 60% of Americans acquire herpes simplex, usually HSV-1, in epithelial lesions commonly known as cold sores. About 30%–60% acquire genital herpes, usually HSV-2, through sexual contact (oral, anal, or vaginal). Genital herpes causes recurrent eruptions of infection in the reproductive tract (**Fig. 11.34**). Many of those infected are unaware of symptoms, but they can still transmit the disease to others.

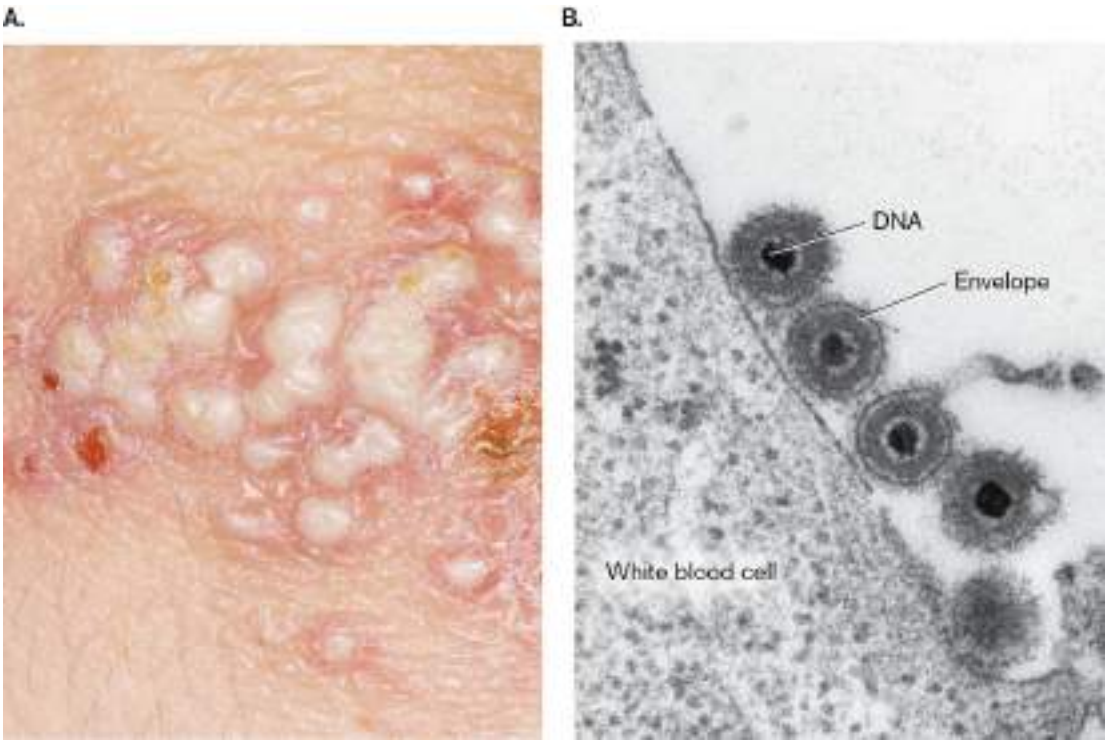


FIGURE 11.34 ■ Genital herpes infection. A. Lesions on the elbow of an 11-year-old patient, caused by HSV-2 infection. **B.** HSV-2 virions, attached to a white blood cell (TEM).

DR P. MARAZZI/SCIENCE SOURCE
DENNIS KUNKEL MICROSCOPY/SCIENCE SOURCE

Herpes simplex virus typically infects cells of the oral or genital mucosa, causing ulcerated sores. The primary infection is epithelial, followed by latent infection within neurons of the ganglia. A common site of infection is the trigeminal ganglion, which processes nerve impulses between the face and eyes and the brain stem.

The latent infection of the ganglia later leads to new outbreaks of virus, often triggered by stress such as menstruation, sunlight exposure, or depression of the immune system. Progeny virions travel back down the dendrites to the epithelia, causing lytic infection. In the trigeminal ganglion, herpes reactivation can lead to eye disease or lethal brain infection. In most cases, herpes symptoms can be controlled by antiviral agents such as acyclovir (discussed in Chapter 27). There is no cure or means of preventing future outbreaks. In pregnant women, HSV can be transmitted to the fetus, with serious complications for the child. HSV can also lead to cases of deadly encephalitis and infectious blindness. eResearch Activity 11 describes an experimental tool of CRISPR-Cas9 delivered by a lentivector to prevent the corneal infection that causes blindness.

Herpes simplex virus is closely related to varicella-zoster virus, the cause of chickenpox, also an epithelial infection. Varicella, too, can hide in ganglial neurons, emerging decades later to cause painful skin lesions called shingles.

Herpes Simplex Virus Structure

Herpes simplex virus has a relatively large genome (more than 70 genes) of double-stranded DNA. The DNA is spooled within an icosahedral capsid under pressure, much like phage lambda, but the herpes capsid is much larger, and it is contained within a host-derived membrane envelope (**Fig. 11.35**).

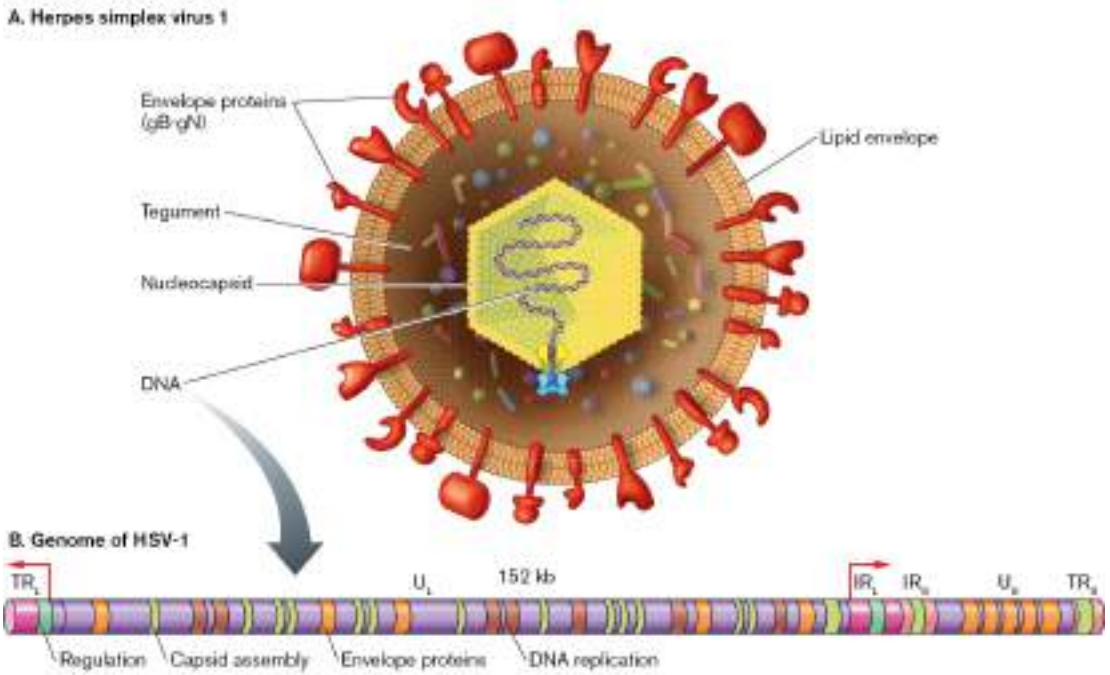


FIGURE 11.35 ■ Herpes simplex virus 1: virion and genome. A. The HSV-1 virion consists of a double-stranded DNA chromosome packaged within an icosahedral capsid. The capsid is surrounded by tegument, a collection of virus-encoded and host-derived proteins. The tegument is contained within a host-derived membrane envelope, including several kinds of envelope proteins. **B.** The genome of HSV-1 spans 152,000 base pairs, encoding more than 70 gene products. The HSV sequence consists of two segments, each containing a unique region (U_L or U_S) flanked by two inverted-repeat regions—terminal (TR_L or TR_S) and internal (IR_L or IR_S)—where the two segments meet.

Within the envelope, the herpes capsid contains double-stranded DNA packaged by inverse spooling (**Fig. 11.35A**). The DNA is packed under pressure to help eject it later, when the uncoated capsid reaches the host cell nucleus. Within the intact virion, the capsid is surrounded by tegument , a collection of about 15 different kinds of virus-encoded proteins, as well as proteins from the previous host. The tegument is contained within a host-derived

membrane envelope with several kinds of spike proteins. The HSV-1 genome spans 152 kb, encoding more than 70 gene products (**Fig. 11.35B**). The sequence includes two unique segments, long (U_L) and short (U_S), each flanked by a terminal repeat (TR_L or TR_S) and an internal repeat (IR_L or IR_S). Within the host, the genome circularizes, so the genetic linkage map appears circular.

The capsid with spooled DNA has a remarkably regular structure, as visualized by cryo-EM of the capsid with envelope removed (**Fig. 11.36A** ; compare with the chapter-opening image). One vertex (corner) of the capsid forms an elaborate portal complex to eject DNA into the host cell nucleus (**Fig. 11.36B**). Inverse spooling generates several layers of wound DNA. At the portal complex, one end of the DNA chromosome is threaded and ready to go, drawing out the entire chromosome with it. Chromosome ejection occurs in the cell when the uncoated capsid docks at a nuclear pore complex.

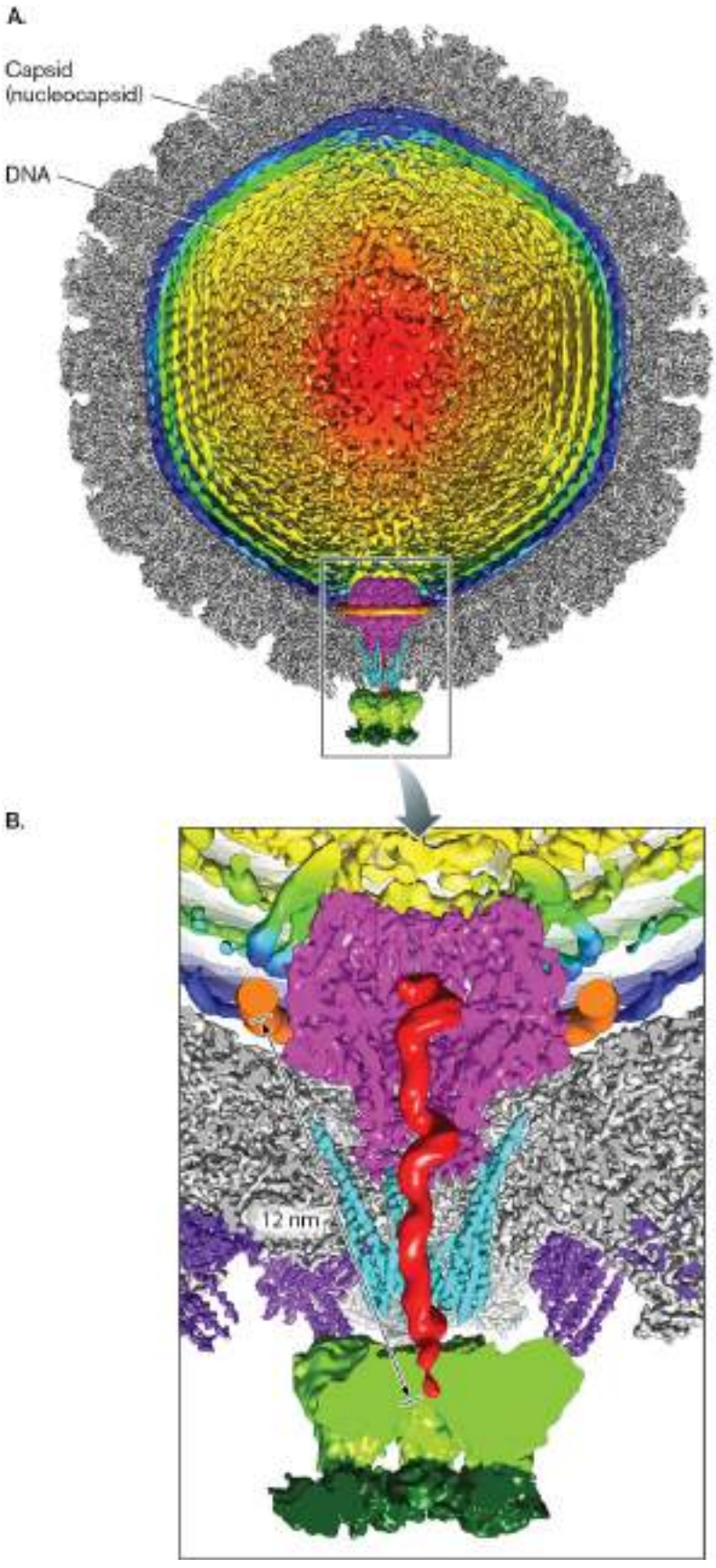

A.
Capsid
(nucleocapsid)
DNA
B.
12 nm

FIGURE 11.36 ■ Herpes simplex virus 1: capsid with DNA portal vertex. A. The HSV-1 capsid contains a double-stranded DNA chromosome spooled within. **B.** One vertex of the icosahedral capsid contains a portal complex for ejection of the DNA.

Y.-T. LIU ET AL. 2019. *NATURE* **570** :257–261

Y.-T. LIU ET AL. 2019. *NATURE* **570** :257–261

Herpesvirus DNA replicates by mechanisms similar to those used in the replication of prokaryotic and phage genomes, either bidirectionally from an origin of replication (as in bacteria) or by the rolling-circle method (as in phages such as T4). As in the replication of cellular genomes, herpesvirus DNA replication requires more than a polymerase; enzymes such as helicase, primase, and single-strand binding proteins are also needed.

Herpes genes and gene regulatory elements resemble those of eukaryotic genes. Each gene has a promoter with eukaryotic control sequences such as the TATA box (see Chapter 8). Genes are transcribed and translated individually; there is no polyprotein. Genes fall into temporal classes within the viral replication cycle: immediate, early, and late expressed genes. Each class has specific regulatory elements. Thus, the virus expresses only the gene products needed for a given phase of infection. An additional class, consisting of the LAT genes, is expressed for latent infection (discussed next).

Herpes Simplex Attachment and Host Cell Entry

Unlike HIV, the relatively large herpes virion has several envelope proteins that can bind to several alternative receptor molecules on the host cell surface, such as a homolog of tumor necrosis factor receptor called HveA or intercellular adhesion molecules called nectins (**Fig. 11.37** , step 1). As with HIV, the entire herpes capsid enters the cytoplasm (step 2). But unlike HIV, whose core particle partly uncoats, the intact herpes capsid travels down a

scaffold of microtubules (step 3) to the nuclear membrane. During this stage, the virion host shutoff factor (Vhs) degrades host mRNA, thus shutting off host protein synthesis. At a nuclear pore complex, the herpes capsid injects its DNA (step 4). The spooled DNA is forced out through the portal complex and into the nucleus by the high pressure of double-stranded DNA packing in the capsid, similar to the high-pressure injection of phage T4 DNA into a bacterial cell. The DNA then circularizes (step 5) to form a plasmid-like intermediate.

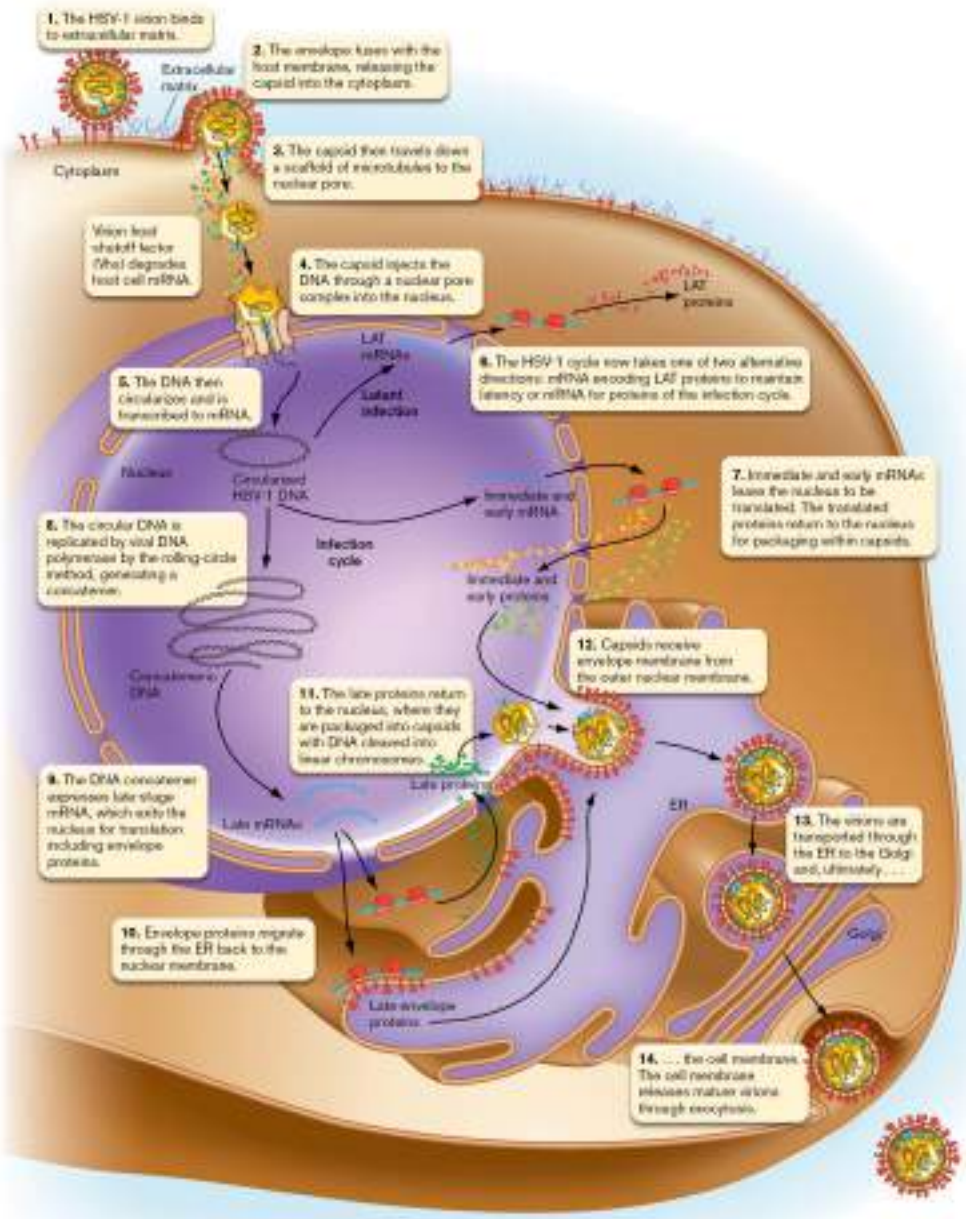


FIGURE 11.37 ■ Replication cycle of HSV-1. The HSV-1 virion binds to receptors on the host cell membrane and releases its capsid in the cytoplasm. The DNA chromosome is transferred into the host nucleus to conduct the replication cycle.

See above for *Herpes Virus Replication* animation

The herpes genome now takes one of two alternative directions: expression of mRNA for proteins of the infection cycle or expression of mRNA encoding LAT proteins to maintain latency (**Fig. 11.37** , step 6). If the latent course is taken, the DNA circle can persist within the cell for decades before switching to lytic infection. Latent infection is seen most often in nerve cells, such as those of the trigeminal ganglion.

Replication of Herpes Simplex Virus

In the nucleus, herpes DNA is transcribed to mRNA by host RNA polymerase II (**Fig. 11.37** , step 5). If mRNA for lytic infection is produced (step 6), it exits the nucleus to be translated by ribosomes. Many different mRNAs are produced and exported, including those required for "immediate" and "early" stages of infection (step 7). The translated proteins return to the nucleus for packaging within capsids.

To generate progeny genomes, the circular DNA is replicated by viral enzymes, including DNA polymerase, single-strand binding protein, and a proofreading endonuclease (**Fig. 11.37** , step 8). Additional enzymes are provided by the host cell. DNA is replicated by the rolling-circle method, generating a concatemer similar to that of phage T4. Unlike T4, however, the herpes DNA is eventually cut into segments defined by the terminal repeat sequences.

The newly synthesized DNA expresses late-stage mRNA (**Fig. 11.37** , step 9), which exits the nucleus for translation. Translated envelope proteins are inserted into the endoplasmic reticulum (ER) membrane, through which they migrate to the nuclear membrane (step 10). Other late proteins reenter the nucleus for assembly into capsids containing DNA genomes (step 11). The viral envelope forms from the outer nuclear membrane (step 12).

The virions are then transported through the ER, where they undergo secondary envelopment (step 13). The secondary-

enveloped virions move to the Golgi, and ultimately to the cell membrane (step 14). The secondary envelope fuses with the cell membrane, releasing mature virions through exocytosis (step 14). Rapid release of virions destroys cells, causing the characteristic sores of herpes infection.

Thought Question

11.9 Compare and contrast the fate of the HSV genome with that of the HIV genome.

Persistent Viral Infections

Herpesviruses can infect humans and other animals indefinitely after initiation of latency by the viral LAT proteins (**Fig. 11.37** , step 6). Most humans are infected by several latent herpesviruses, such as human herpesviruses 6 and 7, Epstein-Barr virus, and cytomegalovirus. Cytomegalovirus is the leading cause of congenital birth defects.



In other respects, however, our bodies may actually benefit from the presence of certain herpesviruses. For example, in 2007 Herbert Virgin and colleagues at the Washington University School of Medicine in St. Louis showed that mice infected with a gamma herpesvirus similar to Epstein-Barr virus resist infection by the bacterial pathogens *Listeria monocytogenes* (the cause of listeriosis) and *Yersinia pestis* (the cause of bubonic plague). The mechanism of antibacterial resistance may involve stimulation of the immune response. Thus, some of our silent herpesvirus "partners" may have coevolved with humans in a mutually beneficial relationship, or mutualism (discussed in Chapter 21).

The maintenance of persistent or latent viral infection involves several kinds of processes that are surprisingly similar in DNA viruses such as herpes and in retroviruses such as HIV. These processes include:

- **Infection of cell types suitable for long-term persistence.** After infecting skin cells for rapid viral replication, some HSV virions infect neurons. Neurons are long-lived cells that provide the virus with an everlasting home in the host. Similarly, Epstein-Barr virus persists within long-lived memory-B-cell lymphocytes.
- **Regulation of viral gene expression.** The LAT proteins suppress expression of viral genes for lytic replication, thus preventing host cell destruction where the latent viral DNA resides. Suppression is the result of assembly of "heterochromatin," chromosome-associated host proteins that inactivate all viral genes except those that encode LAT proteins.
- **Viral subversion of cellular apoptosis.** To maintain latent infection, a viral DNA must prevent host cell apoptosis, a form of programmed self-destruction in response to viral infection. For example, cytomegalovirus prevents apoptosis by mimicking one host apoptosis protein and inhibiting another. Similarly, the HIV retroviral accessory protein Tat prevents apoptosis by inhibiting protein p53, which suppresses tumors through apoptosis.
- **Evasion of immune responses.** Both herpesviruses and retroviruses express proteins that inhibit signaling molecules of the immune system or that mimic immunosuppressive signals.

Herpesviruses Engineered to Kill Tumors

The long-standing adaptation of herpesviruses to the human body confers properties useful for virotherapy. Some kinds of viruses have a preference for infection of tumor cells. Such a virus is called "oncolytic." Oncolytic viruses are being developed to treat cancer. An example of an oncolytic virus for tumor therapy is the engineered

form of HSV-1 called talimogene laherparepvec (T-VEC), developed by a company acquired by Amgen (**Fig. 11.38**). The T-VEC herpesvirus is modified to selectively infect tumor cells of melanoma. In 2015, the FDA approved the use of T-VEC to treat melanoma in patients with inoperable tumors.

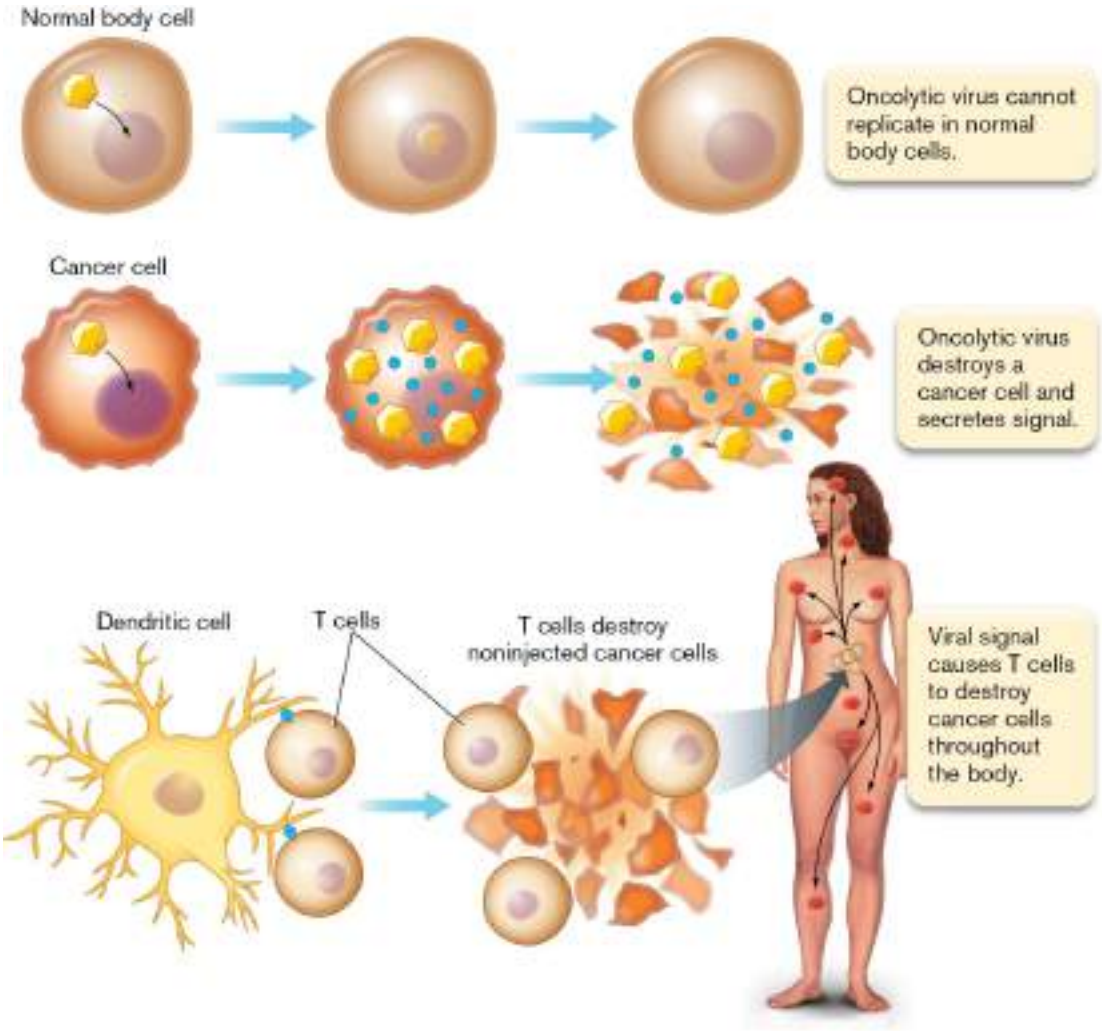


FIGURE 11.38 ■ T-VEC therapy with an oncolytic virus derived from HSV. The oncolytic herpesvirus replicates only in cancer cells, where it destroys the cell and releases signals for host T cells to destroy cancers throughout the body.

The T-VEC virus infects tumor cells and replicates, but it cannot replicate within normal cells. When T-VEC infects tumor cells, it also

releases tumor cell fragments. These tumor fragments are processed by dendritic cells of the immune system (explained in Chapter 24). The dendritic cells present the fragments as antigens to T-cells, which are activated to attack more tumor cells.

Because T-VEC activates the immune system, it helps destroy tumors throughout the body. This antitumor effect of a virus is particularly promising for metastatic tumors that are hard to treat by other means. Besides melanoma, T-VEC is now in trials for treating tumors in advanced cases of other cancers, such as pancreatic cancer and squamous cell cancers of the head and neck.

The study of viral molecular biology raises intriguing questions about the nature of viral replication and about viral origins. Research in virology offers hope for new drugs and cures for humanity's worst plagues, as well as devastating diseases of agricultural plants and animals. At the same time, it is sobering to note that despite the enormous volumes we now know about viruses such as HIV and influenza, the AIDS pandemic continues, and we face new emerging strains of influenza and coronavirus. Molecular research can succeed only in partnership with epidemiology and public health (discussed in Chapter 28).

To Summarize

- **Herpes simplex virus** causes recurring eruptions of sores in the oral or genital mucosa. Initial transmission is by oral or genital contact, followed by eruptions from reactivated virus latent in ganglial neurons.
- **The herpes virion** contains a double-stranded DNA genome packed in an icosahedral capsid. The capsid is surrounded by numerous matrix proteins and by an envelope.
- **HSV attachment** may involve several alternative receptors. A microtubular scaffold transports the herpes virions to the nucleus, where the DNA genome is inserted. The DNA circularizes for transcription.

- **LAT protein expression** leads to latent infection, usually in nerve cells, where the DNA persists silently for months or years.
- **DNA genomes of HSV** are synthesized by the rolling-circle method using viral DNA polymerase supplemented by viral and host-generated components.
- **Infectious mRNA expression** leads to production of capsid, matrix, and envelope proteins for assembly of HSV.
- **HSV assembly** takes place at the nuclear membrane or other membranes. The virions are released from the cell by exocytosis. Rapid release leads to mucosal pathology.
- **HSV can be engineered to specifically kill tumors.** The viral product T-VEC activates the patient's immune system to specifically recognize tumor cells and destroy them.

Glossary

tegument
: The contents of a virion between the capsid and the envelope.

Figure 6.20

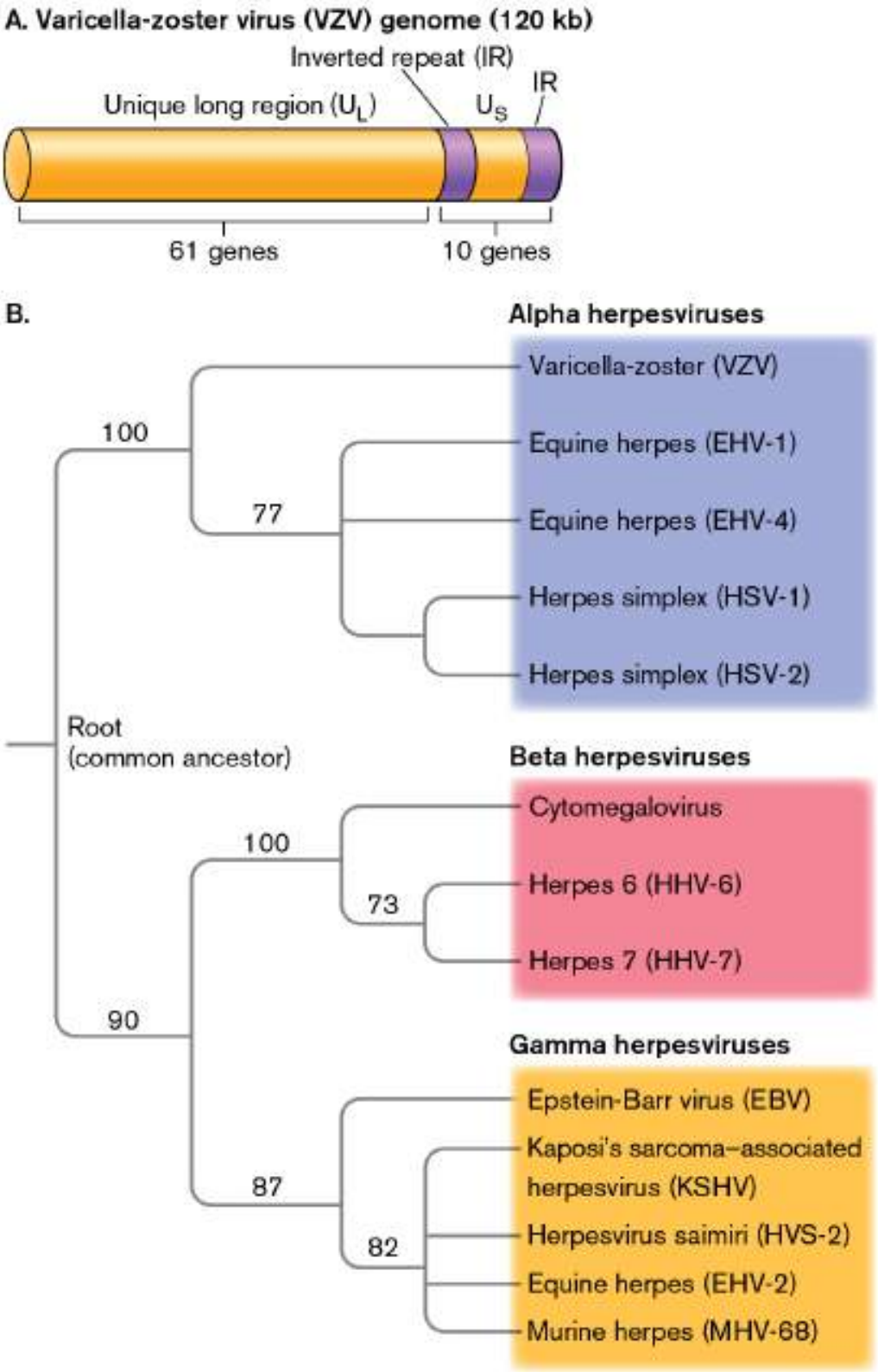


FIGURE 6.20 ■ Phylogeny of herpesvirus genomes. **A.** Genome structure of human varicella-zoster virus (VZV),

the causative agent of chickenpox. **B.** Phylogeny of human and animal herpesviruses, based on whole-genome sequence analysis comparing clusters of orthologous groups of genes. Numbers indicate approximate percentage of sequence shared by divergent genomes from a common ancestor.

eResearch Activity 11

Could a Lentiviral Vector with CRISPR Cure Herpes?

Herpes infections can have devastating effects, and the pathogen always hides out in the ganglia, awaiting the chance for future outbreaks. Could we somehow "erase" the virus from its hiding place? At the Shanghai Center for Systems Biomedicine, funded by the Chinese government, Jiaxu Hong and colleagues are trying to do just that. In a mouse model of corneal infection, they applied a CRISPR-Cas9 guide RNA (described in Chapter 12) carried by a lentiviral vector (**Fig. ERA 11.1**). Thus, a bacterial antiphage host response, plus a lentivector engineered from a lethal retrovirus, could provide a way to remove or suppress herpes simplex virus.

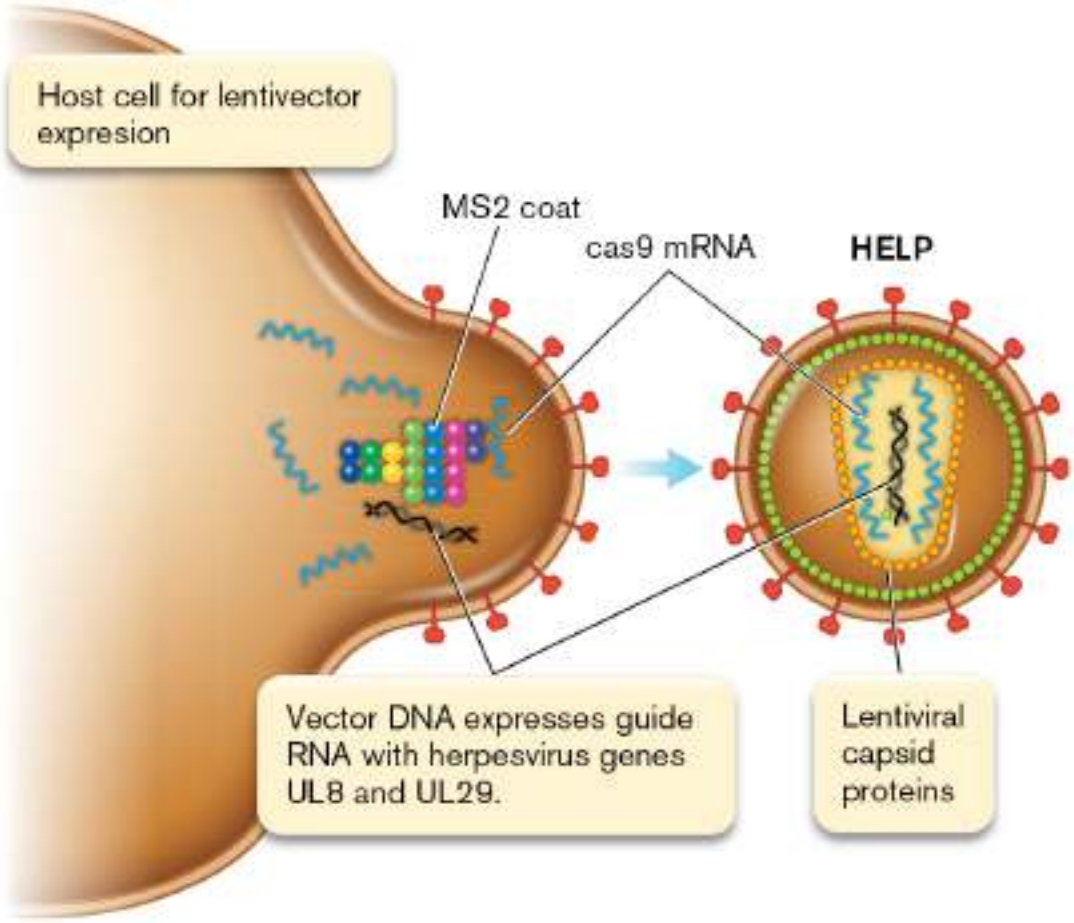


FIGURE ERA 11.1 ■ HELP vector production. Production of HSV-1-erasing lentiviral particles (HELP) using plasmids to carry viral genes essential for infection. HELP particles express a guide RNA that targets herpesvirus genes UL8 and UL29, which are required for viral DNA replication.

The H SV-1- e rasing l entiviral p articles (HELP) are produced by infection of a tissue culture with isolated lentiviral genomes engineered to contain just the genes needed to package the Cas9 RNA for the CRISPR gene-editing reaction. Packaging requires another viral sequence (phage MS2 coat protein) recombined at the N terminus of the Gag and Gag-Pol polyproteins. The MS2 sequence ensures that Cas9 RNA will be included with Gag and Gag-Pol within the lentivector capsid. The lentivector also packages a DNA

expressing two herpesvirus proteins required for DNA replication, UL8 and UL29 (**Fig. ERA 11.1**). Ultimately, the guide RNA will target these two genes in the herpesvirus for cleavage by the Cas9 enzyme.

How is the lentivector packaged without danger of producing infectious lentivirus? The lentiviral proteins essential for infection, such as Gag (capsid) and Pol (reverse transcriptase), are all provided by plasmid expression in the tissue culture. The lentivector will package these products needed for its infection and integration —but its own genetic material contains no genes to express them for progeny virions.

Figure ERA 11.2 shows the results of treating herpesvirus-infected mice with the HELP lentivector. The herpesvirus infects the cornea of the eye, causing an inflammatory disease called keratitis. If untreated, herpes simplex keratitis can lead to blindness.

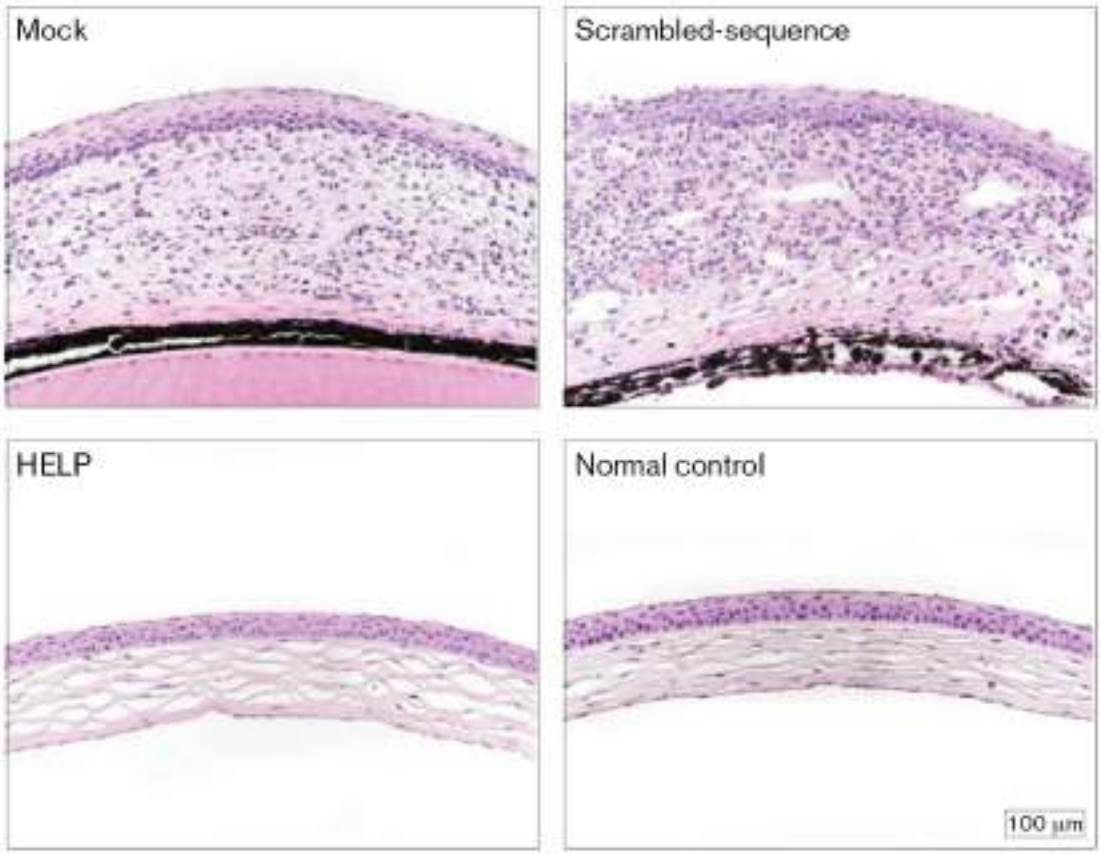


FIGURE ERA 11.2 ■ Herpesvirus-infected mouse corneas treated with HELP. Mock-treated ("Mock") and HELP treatment with scrambled-sequence guide RNA ("Scramble d-sequence") show thickening of the cornea in response to uncontrolled herpes infection. HELP-treated ("HELP") and normal control (no herpes infection) show normal intact cornea.

D. YIN ET AL. 2021. *NAT BIOTECHNOL* . **39** :567–577

Cornea sections stained with hematoxylin-eosin are shown for various experimental conditions. Mock-treated mice (injected carrier buffer, without HELP) show a thickening of the cornea typical of an uncontrolled herpes infection. Similar results appear for mice treated with HELP in which the guide RNA sequence is scrambled, thus preventing targeting of the UL8 and UL29 genes. However, the HELP-treated corneas appear similar to those of the normal control (no herpes infection).

Remarkably, HELP treatment of corneal infection was shown to lead to HELP migration up the nerves to the ganglia, where reservoirs of latent herpesvirus were eliminated. Similar successful results were later reported for a human clinical trial. Thus, in this experiment a construct of multiple viral, phage, and bacterial sources combined to defeat a devastating viral infection—and to remove latent sources of future virus eruption.

Further Exploration

What might be some side reactions of the HELP system, perhaps involving innate immunity? Why is the scrambled guide-RNA control important? How might HELP be re-engineered to treat a viral infection in tissues requiring an altered lentiviral tropism?

Yin, Di, Sikai Ling, Dawei Wang, Yao Dai, Hao Jiang, et al. 2021. Targeting herpes simplex virus with CRISPR–Cas9 cures herpetic stromal keratitis in mice. *Nature Biotechnology* **39** :567.

CHAPTER REVIEW

Review Questions

1. How does phage lambda attach to its correct host cell and insert its genome for replication?
2. How do the CI repressor and Cro protein modulate the switch between lysogeny and lysis?
3. How do influenza virions gain access to the host cytoplasm? Explain the role of fusion peptides.
4. How does influenza virus manage the replication and packaging of its segmented genome? What is the consequence of genome segmentation for virus evolution?
5. Which experimental methods reveal the pathways of intracellular viral replication and identify targets for new antiviral agents?
6. How does HIV provide ready-made components for replication of its genome? How does reverse transcriptase convert the single-stranded RNA genome into double-stranded DNA?
7. What is the role of protease in HIV replication? What is the significance of protease for AIDS therapy?
8. How did ancient retroviruses participate in evolution of the human genome?
9. For gene therapy, how can we construct a lentivector so as to avoid replication of the virus?
10. How does herpesvirus compartmentalize the expression and replication of its DNA genome?

11. Compare and contrast the needs of viral DNA genome replication with those of RNA genome replication.

Thought Questions

1. RNA viruses and DNA viruses represent fundamentally different reproductive strategies. How do their different strategies affect the host response? How can our understanding of viral replication cycles help us develop new antiviral agents?
2. Discuss the roles of host-modulating viral proteins in HIV infection and in herpesvirus infection. What various kinds of functions do these proteins serve for the virus, and what are their effects on the host cell?
3. Hemophilia is a life-threatening genetic blood disorder (the absence of a clotting factor) for which current therapies involve providing blood products and recombinant proteins. Discuss what might be the advantages and relative risks of using viral vectors to treat hemophilia.

Key Terms

- accessory protein (436)
- acquired immunodeficiency syndrome (AIDS) (432)
- azidothymidine (AZT) (439)
- chemokine receptor (CCR) (438)
- concatemer (416)
- core particle (434)
- coreceptor (438)
- endocytosis (427)
- endogenous retrovirus (433 , 444)

- fusion peptide (425)
- gene transfer vector (445)
- host factor (425)
- human immunodeficiency virus (HIV) (432)
- immunohistochemistry (424)
- lentivector (445)
- lentivirus (433)
- long terminal repeat (LTR) (441)
- lysis (418)
- lysogeny (414)
- matrix protein (423)
- nucleocapsid protein (NP) (422)
- prophage (414)
- protease inhibitor (443)
- provirus (441)
- quasispecies (437)
- reassortment (423)
- recombination (423)
- retrotransposon (444)
- retrovirus (432)
- reverse transcriptase (RT) (432 , 439)
- rolling-circle replication (416)
- segmented genome (422)
- spike protein (435)
- tegument (449)
- transfection (447)
- transgene (445)

Recommended Reading

Cartier, Nathalie, Salima Hacein-Bey-Abina, Cynthia C. Bartholomae, Gabor Veres, Manfred Schmidt, et al. 2009.

Hematopoietic stem cell gene therapy with a lentiviral vector in X-linked adrenoleukodystrophy. *Science* **326** :818–823.

Dedrick, Rebekah M., Carlos A. Guerrero-Bustamante, Rebecca A. Garlena, Daniel A. Russell, Katrina Ford, et al. 2019. Engineered bacteriophages for treatment of a patient with a disseminated drug-resistant *Mycobacterium abscessus. Nature Medicine* **25** :730–733.

Deng, Lei, Teena Mohan, Timothy Z. Chang, Gilbert X. Gonzalez, Ye Wang, et al. 2018. Double-layered protein nanoparticles induce broad protection against divergent influenza A viruses. *Nature Communications* **9** :539.

Engelman, Alan, and Peter Cherepanov. 2012. The structural biology of HIV-1: Mechanistic and therapeutic insights. *Nature Reviews. Microbiology* **10** :279–290.

Göke, Jonathan, Xinyi Lu, Yun-Shen Chan, Huck-Hui Ng, Lam-Ha Ly, et al. 2015. Dynamic transcription of distinct classes of endogenous retroviral elements marks specific populations of early human embryonic cells. *Cell Stem Cell* **16** :135–141.

Grow, Edward J., Ryan A. Flynn, Shawn L. Chavez, Nicholas L. Bayless, Mark Wossidlo, et al. 2015. Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. *Nature* **522** :221–225.

Heeney, Jonathan L., Angus G. Dalgleish, and Robin A. Weiss. 2006. Origins of HIV and the evolution of resistance to AIDS. *Science* **313** :462–466.

Joseph, Sarah B., Ronald Swanstrom, Angela D. M. Kashuba, and Myron S. Cohen. 2015. Bottlenecks in HIV-1 transmission: Insights from the study of founder viruses. *Nature Reviews. Microbiology* **13** :414–425.

Karzbrun, Eyal, Alexandra M. Tayar, Vincent Noireaux, and Roy H. Bar-Ziv. 2014. Programmable on-chip DNA compartments as artificial cells. *Science* **345** :829–832.

Klein, Steffen, Mirko Cortese, Sophie L. Winter, Moritz Wachsmuth-Melm, Christopher J. Neufeldt, et al. 2020. SARS-CoV-2 structure and replication characterized by in situ cryo-electron tomography. *Nature Communications* **11** :1–10.

Lakdawala, Seema S., Yicong Wu, Peter Wawrzusin, Juraj Kabat, Andrew J. Broadbent, et al. 2014. Influenza A virus assembly intermediates fuse in the cytoplasm. *PLoS Pathogens* **10** :e1003971.

Liu, Jia, Yufeng Li, Qian Liu, Qun Yao, Xi Wang, et al. 2021. SARS-CoV-2 cell tropism and multiorgan infection. *Cell Discovery* **7** :1–4.

Liu, Qi, Zhe-Min Tan, Jie Sun, Yayi Hou, Congbin Fu, and Zhaohua Wu. 2020. Changing rapid weather variability increases influenza epidemic risk in a warming climate. *Environmental Research Letters* **15** :044004.

Smith, Gavin J. D., Dhanasekaran Vijaykrishna, Justin Bahl, Samantha J. Lycett, Michael Worobey, et al. 2009. Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature* **459** :1122–1126.

Xu, Jun, Roger W. Hendrix, and Robert L. Duda. 2014. Chaperone–protein interactions that mediate assembly of the bacteriophage lambda tail to the correct length. *Journal of Molecular Biology* **426** :1004–1018.

Yosef, Ido, Miriam Manor, Ruth Kiro, and Udi Qimron. 2015. Temperate and lytic bacteriophages programmed to sensitize and kill antibiotic-resistant bacteria. *Proceedings of the National Academy of Sciences USA* **112** :7267–7272.

Zeng, Ming, Zeping Hu, Xiaolei Shi, Xiaohong Li, Xiaoming Zhan, et al. 2014. MAVS, cGAS, and endogenous retroviruses in T-independent B cell responses. *Science* **346** :1486–1492.

Glossary

lysogeny
: A viral life cycle in which the viral genome integrates into and replicates with the host genome but retains the ability to initiate host cell lysis.

prophage
: A phage genome integrated into a host genome.

rolling-circle replication
: A form of DNA replication that proceeds in one direction around a circular template, making tandem copies in a linear array (concatemer). The copies are later cleaved and circularized.

concatemer
: A long line of tandemly repeated genomes; commonly formed during rolling-circle replication.

lysis
: Also called *burst* . The rupture of the cell by a break in the cell wall and membrane.

nucleocapsid protein (NP)
: A protein that coats a viral genome.

segmented genome
: A viral genome that consists of more than one nucleic acid molecule.

reassortment
: The packaging of viral chromosome segments from two different viruses into one progeny virion. Refers to separate segments from a segmented genome, without helix recombination.

recombination
: See also *homologous recombination* . The process by which two DNA molecules exchange arms by cutting and splicing their helix backbones.

matrix protein
: A protein, found in some viruses, that is located between the capsid and the membrane envelope.

immunohistochemistry
: The use of a labeled antibody to stain and visualize specific structures in tissue.

fusion peptide
: A portion of a viral envelope protein that changes shape to facilitate envelope fusion with the host cell membrane.

host factor
: A trait of an individual host that affects susceptibility to disease, in comparison with other individuals.

endocytosis
: The invagination of the cell membrane to form a vesicle that contains extracellular material.

endogenous retrovirus
: A retroelement (genome sequence descended from a retrovirus) that contains *gag* , *env* , and *pol* genes.

reverse transcriptase (RT)
: An enzyme that produces a double-stranded DNA molecule from a single-stranded RNA template.

human immunodeficiency virus (HIV)
: A human-specific retrovirus that causes AIDS.

acquired immunodeficiency syndrome (AIDS)
: A disease caused by HIV that leads to the destruction of T cells and the inability to fight off opportunistic infections.

lentivirus
: A member of a family of retroviruses with a long incubation period. An example is HIV.

core particle
: A viral capsid that encloses its nucleic acid genome and is surrounded by an envelope.

spike protein
: A viral glycoprotein that connects the membrane to the capsid or the matrix and may be involved in viral binding to host cell receptors.

accessory protein
: A protein found in the viral capsid or tegument that is needed early in the viral life cycle.

quasispecies
: A collection of isolates (usually viruses) from a common source of infection that have evolved into many different types within

one host.

chemokine receptor (CCR)
: A human T-cell membrane protein that binds chemokine hormones but is also used by HIV for attachment and infection.

coreceptor
: A cell-surface receptor needed for viral entry along with a primary receptor.

azidothymidine (AZT)
: A nucleotide analog that inhibits reverse transcriptase and was the first drug clinically used to fight HIV infections.

provirus
: A viral genome that is integrated into the host cell genome.

long terminal repeat (LTR)
: A repeated nucleic acid sequence at the 5′ and 3′ ends of a provirus.

protease inhibitor
: A molecule that inhibits a protease enzyme; some are used as anti-HIV drugs to block the virally encoded protease needed to complete HIV assembly.

retrotransposon
: A retroelement that contains only partial retroviral sequences but may encode reverse transcriptase to allow further movement into the host genome.

gene transfer vector
: A mobile DNA engineered from a virus or plasmid, designed to insert a genetic sequence into the genome of an organism for experimental study or for medical therapy.

lentivector *or* lentiviral vector
: A gene transfer vector derived from a lentivirus such as HIV; designed to integrate genes into a host chromosome.

transgene
: A gene that has been transferred by genetic engineering techniques from one organism to another.

transfection
: In biotechnology, the transfer of DNA (usually viral) into cells.

tegument

The contents of a virion between the capsid and the envelope.

retrovirus

Also called *RNA reverse-transcribing virus* . A single-stranded RNA virus that uses reverse transcriptase to generate a double-stranded DNA.